In Anna Esposito · Marcos Faundez-Zanuy Antonietta M. Esposito · Gennaro Cordasco
Thomas Drugman · Jordi Solé-Casals Francesco Carlo Morabito (Eds)
Recent Advances in Nonlinear Speech Processing. (p. 15-34)

The Relationship Between the (Mis)-Parsing of Coarticulation in Perception and Sound Change: Evidence from Dissimilation and Language Acquisition

Jonathan Harrington, Felicitas Kleber and Mary Stevens

Abstract The study is concerned with whether historical sound change is more likely to occur when coarticulation, or the way that speech sounds overlap with and influence each other in time, is misaligned in production and perception. The focus of the first experiment was on long-range coarticulatory lip-rounding that has been linked with historical dissimilation. A perception experiment based on present-day Italian showed that inherently lip-rounded segments were more likely to be masked—and thereby erroneously deleted—in hypoarticulated speech. The second experiment tested whether the mismatch between the modalities was more likely in young children than in adults. For this purpose, first language German speakers participated in a forced-choice perception experiment in which they categorised German back and front vowels in coarticulatory non-fronting and fronting consonantal contexts. Children's ability to normalise for coarticulation was shown to be less than that of the adults. Taken together, the results suggest that sound change can occur when coarticulatory relationships are perceptually obscured due to a hypoarticulated speaking style causing consonants to be camouflaged in the case of dissimilation and variants to approximate those that are strongly influenced by coarticulation in the case of diachronic back vowel fronting.

1 Introduction

Research in the last 30–40 years has shown a relationship between contextual variation in speech communication and historical change. A well-known example is synchronic transconsonantal vowel coarticulation [24, 52] that has led to the sound

J. Harrington (✉) · F. Kleber · M. Stevens
Institute of Phonetics and Speech Processing, Ludwig-Maximilians
University of Munich, Munich, Germany
e-mail: jmh@phonetik.uni-muenchen.de

F. Kleber
e-mail: kleber@phonetik.uni-muenchen.de

M. Stevens
e-mail: mes@phonetik.uni-muenchen.de

A. Esposito et al. (eds.), *Recent Advances in Nonlinear Speech Processing*,
Smart Innovation, Systems and Technologies 48,
DOI 10.1007/978-3-319-28109-4_3

change by which umlaut has developed in some languages (e.g. present-day German *Füße* /fyse/ and present-day English ‘feet’ from Proto-Germanic /fotiz/). The general aim in this chapter is to consider the mechanisms by which diachronic change can take hold, given what is known about the dynamics of speech production and their relationship to perception. The focus will be on the types of sound change that have been documented in numerous languages and whose bases are in coarticulation: that is in how speech sounds overlap with and influence each other in time.

Many physiological, acoustic, and perceptual studies are consistent with the idea developed from action theory [15] through to articulatory phonology [7] that speech production can be modelled as the orchestration of autonomous gestures that wax and wane in time [45] so that, within any given time window multiple sounds make contributions in different degrees of strength to the acoustic signal [59]. Thus, in producing *queen* /kwin/, the tongue-dorsum raising of /i/ is likely to overlap partially or entirely both with the preceding lip-rounding from /w/ and with a lowered velum in anticipation of the following /n/. This simultaneous production, coarticulation or coproduction of multiple gestures from successive speech sounds can also easily extend across major prosodic boundaries, especially in the case of liquids [23, 32, 61]. The important point as far as modelling sound change from coarticulation is concerned is that, to use an apt metaphor from Lindblom [43], speech is ‘big band’ in which the gestures of speech production are independently controlled and each make their own contribution to the acoustic signal, analogously to the acoustic contribution towards realising a common musical arrangement that is made by the independently controlled and separate instruments of an orchestra.

As far as the listener is concerned, numerous experiments suggest that coarticulation is perceived analogously to its production. The evidence for this derives from experiments showing how an identical acoustical signal is differently perceived depending on the context in which it is embedded [21, 46]. For example, many listeners hear a nasal vowel as *oral* when it is surrounded by nasal consonants [34]. An explanation for this finding is that listeners parse the acoustic signal into the overlapping articulatory gestures that could have given rise to it [17, 20]. For this example, listeners perceive the temporally overlapping tongue dorsum movements for the vowel and lowered velum of the nasal as autonomously coproduced gestures. A consequence of perceived coproduction is that nasalisation in perception is associated not with the vowel but with the nasal consonant that caused it: that is, listeners factor nasalisation from the acoustic signal of the vowel [4, 5]. It is in this sense that some have argued for parity or a common currency between coarticulation in production and coarticulation in perception [16]. For some researchers, there must necessarily be parity because gestures are directly perceived in the speech signal [19]. In the theory to be developed in the present chapter, such parity across production and perception in processing coarticulation is considered to be the condition that obtains only under stability, i.e. when no sound change is taking place. Moreover, we propose that sound change can occur under the perhaps rarer condition in which the production and perception of coarticulation are out of alignment: that is, when listeners perceive or parse coarticulation in a way that is different from its production.

This theory is closely informed by Ohala's [48–50] model of sound change and extensions thereof (e.g. [41, 57]) in which occasional ambiguities in the transmission of coarticulation from a speaker to hearer can be a source of sound change. A well-known example is the epenthetic stop insertion that synchronically gives rise to variations such as /drɛmt, drɛmpt/ ('dreamed') that are related to sound changes such as *empty* from old English /æmtiɡ/. Where then does the /p/ come from? A likely answer is that a /p/ can be perceived if the lip closure for /m/ is released not synchronously but after the oral closure leading to a bilabial stop or doubly articulated [p͡t]. Notice that the listener must have heard a /p/ even though no such unit formed part of speech production. That this must be so can be seen in the derivation of names such as 'Hampton' which arose by combining in Old English the surname 'Ham' (and importantly not 'Hamp') with 'tun'. Thus, the part of the signal corresponding to the overlapping lip-constriction and /t/ closure that must have occurred in /æmtiɡ/ and /hamtun/ has been decontextualised by the listener, because it is not interpreted in relation to the phonetic context that gave rise to it. This part of the signal has (with the passage of time) instead been *phonologised* because it has come to be permanently associated (in 'empty' and 'Hampton') with a /p/ phoneme where none had originally existed.

Speech communication also varies substantially in speaking style. This can be in a social sense, as when speakers adapt their style to take account of the social status of the interlocutor (e.g. [30]). Moreover, there is, of course, well-documented evidence of an association between the adaptation of speaking style towards a more prestigious social class and sound change [38]. Here we shall be concerned not with social variation, but instead with the adaptation in speech production depending on the extent to which the meaning of the signal is predictable from context. According to Lindblom [42], speech is produced with a high degree of clarity or *hyper*articulation when the listener has to rely almost entirely on the signal to understand it. This might happen in introducing a person for the first time, given that there is unlikely to be any prior context or knowledge by which the listener can infer the person's name from context. Local hyperarticulation is likely to occur at points in the signal that are particularly important for understanding what is being said [11]; in stress-accent languages, these points of information focus also typically occur in nuclear accented words [10]. By contrast, a speaker tends to *hypo*articulate the parts of the speech signal in which the listener is predicted to be able to bring to bear contextual knowledge in the broadest sense—sometimes because of a topic that is current in a dialogue, sometimes by means of the knowledge that is assumed to be shared by the speaker and listener [54]. According to Lindblom [42], listeners tend not to process the details of the signal in hypoarticulated speech: firstly, they might not need to because the signal should be highly predictable using top-down information; secondly, hypoarticulated signals may in any case be of less use for decoding meaning if the phonetic content is degraded—such as when vowels are reduced and consonants are strongly lenited, as is typical of a hypoarticulated speaking style. In Lindblom [44], it is when listeners exceptionally process the fine phonetic detail in hypoarticulated speech that a new pronunciation for a word can be added to the lexicon.

Harrington et al. [28, 29] tested whether the types of ambiguities in the transmission of coarticulation—that Ohala considers to be responsible for many kinds of sound change—may be more acute in hypoarticulated speech. They assessed whether prosodic weakening influenced the extent to which listeners adjusted their perceptions for a coarticulatory effect (vowel fronting in VCV coarticulation and polysyllabic shortening). Their results suggested less perceptual adjustment for coarticulation in lexically weak than strong syllables [28] and less adjustment in prosodically deaccented than accented words [29]. Taken together, these results provide some evidence that listeners' phonological categorisations are less influenced by coarticulation in hypoarticulated speech (of which lexically weak syllables and deaccented words are two examples). In this paper, we extend this idea to test whether there is a connection between dissimilation sound changes and the degradation of perceived coarticulation in hypoarticulated speech (see [2] and [6] for a review and analysis of dissimilation in different languages). Dissimilation is a very different type of sound change compared with those that formed the basis of the analysis in Harrington et al. [28, 29] in which the sound changes associated with phonetic variation come about because listeners are presumed to adjust their perceptions insufficiently for coarticulation. Dissimilation, by contrast, comes about according to Ohala [53] because listeners adjust their perceptions *too much* for a presumed coarticulatory effect. So far, there have been very few attempts to reconstruct in the laboratory the synchronic conditions that could lead to dissimilation and the few that have been conducted (e.g. [1]) have found little evidence to support the idea that dissimilation is associated with an over-compensation for coarticulation, as suggested by Ohala [48, 49].

In the last 10 years, various studies have shown that listeners even of the same dialect do not always agree on how to process coarticulation. For example, studies by Beddor [4, 5] have shown that American English listeners vary in how nasalised they perceive a vowel to be before a nasal consonant. Moreover, Yu [63] and Yu et al. [64] have shown how listeners' perception of (and normalisation for) coarticulation is influenced by their personality and social profiles. Such results suggest another potential source of sound change: that individuals or perhaps groups of individuals differ in how a given speech signal is parsed perceptually (see also [33]). Group differences in processing coarticulation perceptually were found for older versus younger subjects for a sound change in progress in Standard Southern British [27, 36].

In this chapter, we consider whether the possibly different ways in which adults and young children perceive coarticulation may be another potential source of sound change. There is of course an extensive literature on the association between sound change and language acquisition [22, 35, 40] with a particular emphasis on demonstrating the commonality between children's misarticulations during acquisition and patterns of sound change. As argued elsewhere [3, 14, 60], there is little evidence for such a direct association and this is also not the type of investigation that is being pursued here. The approach follows instead that of Kleber and Peters [37] who seek to test whether, as less experienced users of the language, children are more likely to have difficulty normalising in perception for coarticulation. Some evidence that this might be so was presented in Nittrouer and Studdert-Kennedy [47] for consonants.

Since that study, there have been no further analyses of whether adults and children process coarticulation differently in perception. Here we extend their analysis for the first time to an investigation of the coarticulatory influences of consonants on vowels (and in another language).

In summary, the aim of this chapter is to test earlier [50] and more recent [4, 36, 41, 57] models of an association between sound change and the perceptual processing of coarticulation. We approach this issue from two very different perspectives. Firstly, by considering how listeners' processing ambiguities could give rise to dissimilation (Sect. 2). Secondly, by analysing whether sound change might arise through the different ways that coarticulation might be processed across two groups of listeners—in this case children and adults (Sect. 3).

2 Sound Change and Dissimilation

The experiment in this section was concerned with the relationship between perceptual processing and a dissimilatory sound change. Dissimilation occurs when one of two similar segments in close proximity changes to become less similar. An example is Grassman's Law under which aspiration disappears when there is another following aspirated stop, e.g. Ancient Greek /t**ʰ**riks/ 'hair' nominative' but /trik**ʰ**os/ 'hair' (genitive) derived historically from /t**ʰ**rikʰos/ with initial aspiration. According to Ohala [50], dissimilation can occur when listeners mistakenly attribute part of a speech sound to coarticulation instead of to the speech sound itself. For the example above, sound change comes about because listeners mistakenly interpret the first aspirated segment as being caused by anticipatory coarticulatory spreading of the second /h/. A similar idea is used to explain the sound change whereby the first /w/ was deleted from Latin /k**w**ink**w**e/ ('five') leading to /kink**w**e/ (and then via a different sound change to /tʃink**w**e/ in present-day Italian). The interpretation in Ohala's model is that there is long-range lip-rounding between the two /w/s in /kwinkwe/ that listeners attribute to the second /w/. From another point of view, long-range lip-rounding due to coarticulation occasionally prevents listeners from interpreting the first /w/ as a phonological unit in its own right. Notice that this sound change did not apply to Latin *quindecim* ('fifteen') which is produced with an initial /kw/ in present-day Italian, i.e. /kwinditʃi/). This is because there is no /w/ that occurs later in this word to which coarticulation could be incorrectly attributed.

In the following experiment, two hypotheses were tested. The first was concerned with creating the conditions in the laboratory that could have given rise to dissimilation by testing whether, in present-day Italian, a later occurring /w/ could mask the perception of an initial /w/. Since there are very few words in Italian with a repeated /w/ (*qualunque* being one of the few exceptions), this was done by testing the perception of /kw/ versus /k/ in a target word when followed by a word that did (*quattro*, 'four', /kwat:ro/) or did not (*sette*, 'seven', /set:e/) contain a prevocalic /w/. The second hypothesis was concerned with the effects of hypoarticulation. Here we tested whether the perception of the first /w/ was even more likely to be masked by coarticulation when the target word occurred in a deaccenting/hypoarticulation context. The reasoning behind this follows the arguments of the preceding section that the perceptual parsing of coarticulation may be obscured in hypoarticulated signals: that is, hypoarticulated speech may blur the distinction between lip-rounding due to the presence of an /w/ and the long-range coarticulatory effects of lip-rounding that arise due to the second /w/.

2.1 Method

We created a /kw…k/ continuum and tested the effect of the following word (an initial /kw/ vs. /s/) and prosodic context (accented or deaccented) on listener perception of lip rounding. To create the stimuli, we extracted a single *canto* (/kanto/ 'I sing') token from phrases that had been read aloud and recorded by a female Italian speaker for the purposes of this experiment. We used PSOLA in Praat to synthesize an 11-step continuum from /kanto/ to /kwanto/ (*quanto*, 'how much'). We also lowered F2 in /anto/ to simulate lip rounding throughout the word. We discarded steps 2 and 10 to keep the time taken for the experiment as short as possible for participants. The resulting 9-step *canto…quanto* continuum was inserted into four different carrier phrases that differed according to the following word (*quattro* vs. *sette*) and to whether or not the target word *canto-quanto* was accented (shown in upper case below) or deaccented. In the accented condition, the nuclear accent fell on *canto-quanto* which was synthesised with a large pitch obtrusion appropriate for an L+H* pitch-accent on the first syllable /kan/. In the deaccented condition, the nuclear accent fell on *detto* ('said'), the pitch obtrusion occurred on the first syllable of 'detto' and *canto-quanto* were deaccented (synthesised with a low and flat pitch). The two readings differ in the location of (narrow) focus: thus, the accented condition might be appropriate as a response to 'what did you say four/seven times?' and the deaccented condition as a response to 'did you read or did you say *canto-quanto* four/seven times?' (see [12] and [39] for further details on the association between focus and accent in Italian).

$$
\begin{array}{lllll}
H0 \left\{ \begin{array}{l} \textit{detto QUANTO}\ldots\textit{CANTO} \\ \textit{DETTO quanto}\ldots\textit{canto} \end{array} \right\} & \times & \left\{ \begin{array}{l} \textit{quattro} \\ \textit{sette} \end{array} \right\} & \textit{volte} & \begin{array}{l} \text{`I said\_\_\_four times'} \\ \text{`I SAID\_\_seven times'} \end{array} \\
\quad\text{Prosodic context} & & \text{Following word} & &
\end{array}
$$

The stimuli (9 continuum steps × 2 prosodic contexts (accented/deaccented) × 2 following words (*quattro*/*sette*) × 10 repetitions = 360) were presented to 24 Italian listeners in a two-alternative forced-choice perception test that was conducted online. Participants also heard additional stimuli consisting of the target words *canto-quanto* in isolation; we do not discuss the isolated word data here. Participants were asked to wear headphones and could listen to the stimuli as many times as they wished. Their task was to listen to each phrase stimulus, decide whether the target word sounded more like *canto* or *quanto* and click on the corresponding button. The listener participants were native Italian speakers aged between 19 and 53 years and, in terms of regional variety, all but two were self-reported Standard Italian and/or Tuscan Italian speakers. All participants were paid for their participation with a voucher sent to their email address.

We fitted a generalized linear model within the R package `lme4` with the listener response (2 levels: *quanto*/*canto*) as the dependent variable, prosody (2 levels: accented/deaccented), word (2 levels: *quattro*/*sette*) and the stimulus number as fixed factors, and also included all two-way interactions between these factors. The listener (24 levels) was included as random factor. The significance of any term was obtained by testing whether the full model and one without the term being tested differed significantly from each other.

2.2 Results

We excluded 3/24 listeners from all further analyses because there was no convergence in their derived psychometric curves (i.e. the decision boundaries for these three listeners lay well beyond the range of the stimulus steps).

Our first hypothesis is that the following word (*quattro* vs. *sette*) should influence listeners' decisions and that there should be more *canto* responses when the target word precedes *quattro*. This is because, following Ohala's model, listeners should attribute lip rounding during the target word to anticipatory coarticulation for the upcoming /kw/. But there should be no such bias towards *canto* in the *sette* context since there is no /w/ in the following word to which coarticulation could be attributed. It is clear from Fig. 1—which shows psychometric curves fitted to all 21 listeners separately in the four contexts—that there is no support for this hypothesis. If there had been more *canto* responses preceding *quattro*, then the black curves in Fig. 1 should be to the left of the grey ones. In fact, there appear to be no differences in responses before the two words in the deaccented context while in the accented context, there is even a trend towards more *quanto* responses preceding *quattro*. Thus there is no evidence from this experiment to support the idea that a following /w/ masks the perception of an initial /w/.

The second hypothesis was that there should be an even greater tendency for more *canto* responses before *quattro* in the deaccented condition. Obviously, there is no support for this hypothesis either, given the completely overlapping psychometric curves preceding these words in the deaccented context. On the other hand, the same figure shows that there are very many more *canto* responses in the deaccented than in the accented context. Some suggestions for why this might be so are discussed below.

The observations in Fig. 1 were to a large extent supported by the statistical analysis. Firstly, dropping word as well as both the interaction of word with stimulus and of word with prosody made no significant difference to the statistical model. This result shows that the difference between *quattro*/*sette* had no significant influence on the responses: thus the trend in Fig. 1 by which there were more *quanto* responses pre-

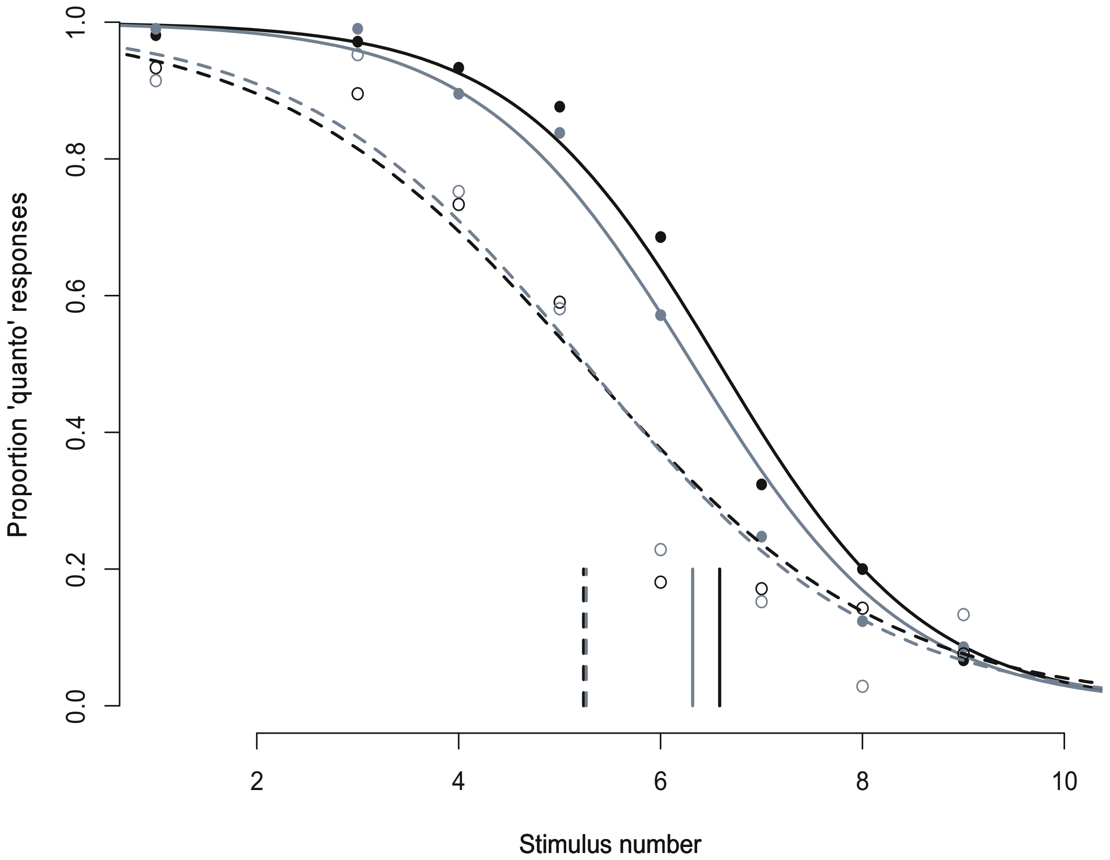

Fig. 1 Fitted psychometric curves showing listener responses to a *quanto-canto* continuum embedded in four contexts: Preceding __*quattro* (*black*) versus __*sette* (*grey*) and in accented (*solid*) versus deaccented (*dashed*) position. The *circles* are the averages of responses across all listeners for any stimulus number. Increasing stimulus numbers are from low to high F2

ceding *quattro* in the accented condition was not supported by the statistical analysis. Consistently with Fig. 1, there was a significant interaction between stimulus number and prosody ($\chi^2_1 = 31.2$, $p < 0.001$): this result shows what is evident in Fig. 1 that there are more *canto* responses along the stimulus continuum in the deaccented versus accented conditions.

2.3 Discussion

We simulated long-range coarticulatory lip-rounding both by lowering F2 in—*anto* and through the presence of /w/ in an immediately following word. The presence of a lip-rounded consonant in the following word (*quattro*) or not (*sette*) made no difference to listeners' responses. We were therefore not able to recreate in the laboratory the sound change by which an initial /w/ in Latin *quinque* dissimilates as a result of a following /w/. This negative finding does not necessarily mean that Ohala's [50] idea about dissimilation through over-compensating for coarticulation is wrong; it may just be that these following word differences are insufficient for recreating in an experiment the conditions by which historic dissimilation could have occurred. But there is, however, an interpretation that could be consistent with Ohala's model based on our second finding that, regardless of the following word, listeners responded far more with *canto* when the target word was deaccented. Our interpretation of this finding is that in deaccented/hypoarticulated speech, the separation between long-range anticipatory lip-rounding from lip-rounding due to the initial /w/ is obscured. Recall that in our stimuli, -*anto* was in all cases synthesised with a very low F2. This makes all the stimuli sound as if they were produced by a speaker with a long-term rounded lip setting. It is this speaker-attribute of lip-rounding that listeners fail to distinguish perceptually from the /w/ of *quanto* (causing them to hear *canto*). Thus the lip-rounding in our data camouflages perceptually the initial /w/. This is entirely consistent with Ohala [48] who also interprets dissimilation as perceptual camouflage.

According to Ohala, the listener error that is the source of dissimilation comes about because of a following /w/. We were not able to demonstrate that with our results. But our results are consistent with the idea that long-range lip-rounding can interfere with the perception of an initial /w/. The further new angle suggested by the present results is that this interference comes about not in all speaking styles, but specifically in hypoarticulated/deaccented speech.

The present study and those in Harrington et al. [28, 29] have suggested a language-internal motivation for sound change which arises because hypoarticulation (simulated here by deaccenting) can be detrimental to parsing coarticulation perceptually. In the next experiment, we consider the extent to which differences at the group level—between adults and children in parsing coarticulation—may additionally contribute to some of the conditions that can cause sound change to occur.

3 The Perception of Coarticulation by Adults and by Children

The coarticulation to be investigated in this experiment was the fronting of the mid-high lax rounded vowel /ʊ/ in a symmetrical /t_t/ context and the acoustic lowering of the mid-high front rounded vowel /ʏ/ in a symmetrical /p_p/ context. The materials were in all cases taken from standard German in which /ʊ, ʏ/ are contrastive (e.g. *musste*/*müsste*, /mʊste, mʏste/, ‘had to’/‘should have’).

The coarticulatory fronting of /ʊ/ has been extensively documented and comes about because the tongue dorsum for /ʊ/ is shifted forward under the influence of the alveolar constriction [36, 52]. This type of phonetic /ʊ/-fronting causes a raising of its second formant frequency. The coarticulatory F2-lowering in /ʏ/ comes about because the constricted lip gesture for /p/ overlaps with the high front rounded vowel /ʏ/. From another related point of view, the F2-frequency of /ʏ/ is lowered under the influence of the low F2-locus frequency for /p/.

The influence of coarticulation in perception was tested using a well-established technique of embedding an acoustically identical /ʊ-ʏ/ continuum in /t_t/ and /p_p/ contexts and then deriving through a forced-choice listening experiment the cross-over boundary at which responses are equivocal, i.e. at 50 % [21, 46]. The main point to observe here is that the direction in which /ʊ, ʏ/ differ acoustically (from low to high F2) is the same as that of the coarticulatory influence of the /p_p, t_t/ contexts (also from low F2 for /p_p/ to high F2 for /t_t/). Therefore, if listeners adjust their responses in relation to these coarticulatory effects, then they should be more likely to hear /ʊ/ in a /t_t/ than a /p_p/ context. This is further illustrated in Fig. 2 which shows schematically the relationship between production and perception for these vowel × context combinations. The figure shows how the distributions are shifted to the right (towards higher F2) in the production of both vowels in the /t_t/ context than in a /p_p/ context for the reasons stated above: /t/ causes F2 to be raised, and /p/ F2 to be lowered (with the raising effect due to the alveolar possibly being greater than the lowering effect due to the labial). Consequently, if perceptions are adjusted exactly for these effects of coarticulation—that is, if there is 'parity' between the production and perception of coarticulation—then the cross-over boundary in perception from /ʊ/ to /ʏ/ (shown in the lower half of the same figure) should be higher in a /t_t/ than in a /p_p/ context. The issue to be tested is whether this difference in the cross-over boundary (the length of the line marked 'normalise' in Fig. 2) was less for children, which would indicate that they normalise less for coarticulation. This follows from

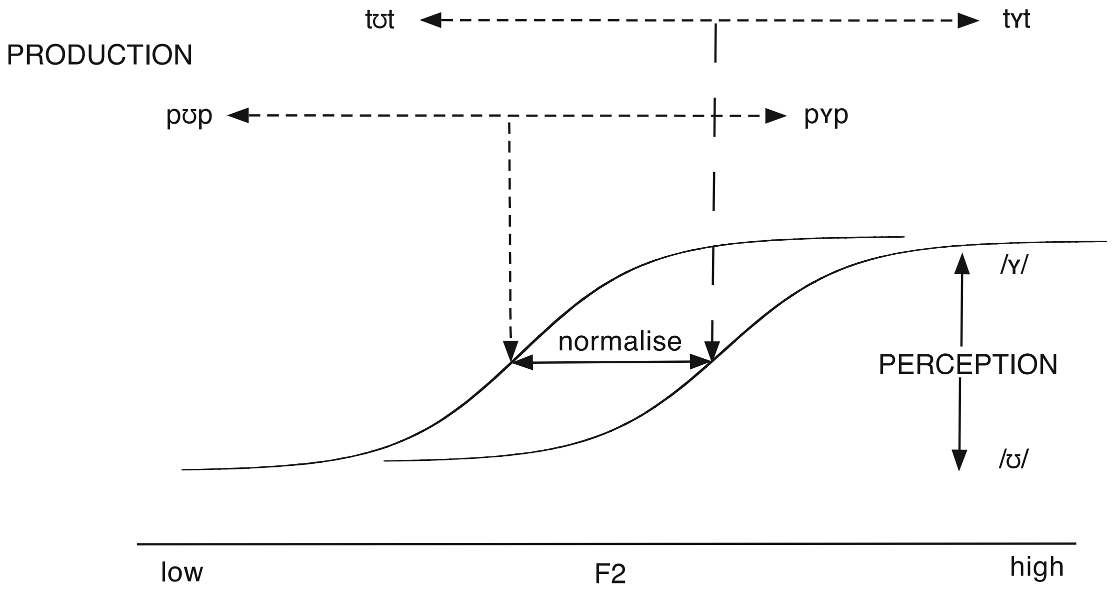

Fig. 2 A schematic outline of the relationship between the production and perception of coarticulation. The *upper part* of the display illustrates the hypothetical distributions in F2 of the four target words showing higher F2 values in a /t_t/ context. The *lower panel* shows the distribution of the corresponding perceptual responses under the assumption that the production and perception of coarticulation are exactly aligned. The degree to which listeners normalise for coarticulation in this model is proportional to the length of the horizontal line marked 'normalise' which extends between the two sigmoids cross-over boundaries at which the probability of perceiving /ʊ/ or /ʏ/ are both 50 %

one of Ohala's [50] predictions that children as less experienced listeners of the language might normalise less for coarticulation than adults.

3.1 Method

There were three parts to the method. Firstly (Sect. 3.1.1), a training phase for the children which also involved the creation of a child-production database. Secondly (Sect. 3.1.2), the creation of perception stimuli to which we obtained forced-choice categorical responses from the adults and imitations (following training) from the children. Thirdly (Sect. 3.1.3), the conversion of the child imitations to categorical responses. These three stages are described more fully below.

3.1.1 Child-Production Database

13 children participated in a training period in which they first learned to associate the target non-words with four puppet names TUTT, TÜTT, PUPP, PÜPP corresponding phonemically to /tʊt, tʏt pʊp, pʏp/ respectively (Fig. 3). Once these had been learned, they produced each of the four puppet names five times. The productions were obtained from the children from a randomised sequence of the puppets' pictures (those in Fig. 3) that were presented on a computer screen one at a time. This child-production database consisted of 4 words × 5 repetitions × 13 children = 260 tokens.

3.1.2 Creation of Synthetic Stimuli

A male speaker of Standard German with slight South German regional characteristics produced utterances containing /pʊp, tʏt/ in the carrier phrase 'Maria hat ___ gesagt' (literally: 'Maria has ___ said') with nuclear accent on the target word. An

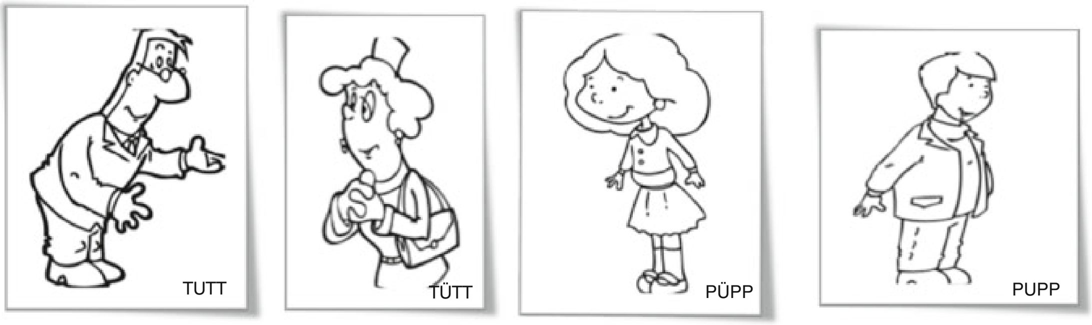

Fig. 3 The four puppet pictures used in the picture-naming task by the children for the production (from *left* to *right*) of /tʊt, tʏt, pʏp, pʊp/

11-step F2 continuum was created between original productions of /pʊp/ (F2 = 803 Hz) and /tʏt/ (F2 = 1436 Hz) by using LPC-resynthesis in the static morphing method of Akustyk [56]. The durations of the vowels were normalized using PSOLA. This same 11-step vowel continuum differing in F2 was spliced into labial /p_p/ and alveolar /t_t/ contexts in the same utterance ‘Maria hat ___ gesagt’ with nuclear accent on the target word (see also [36] for further details).

The stimuli were randomised and presented to a group of 20 L1-German speaking adults (students at the IPS, most of them in their twenties) and to a group of 13 L1-German speaking children (age range from 4 years and 11 months to 6 years and 3 months) resulting in 2 continua (p_p, t_t) × 11 stimuli × 10 repetitions × 20 adult listeners + 2 continua (p_p, t_t) × 11 stimuli × 3 repetitions × 13 children listeners = 5258 presentations. The adults carried out a forced choice identification task and identified each stimulus as one of TUTT, TÜTT, PUPP, PÜPP. Since such a task was considered to be too difficult for the children, they instead imitated each stimulus that they heard following both the training period and the creation of the child-production database as described in Sect. 3.1.1: that is, the children were very familiar with the four characters shown in Fig. 4 before they participated in this perception and imitation experiment.

3.1.3 Obtaining Categorical Responses from Children

Both the child-production (Sect. 3.1.1) and child-imitation (Sect. 3.1.2) corpora were segmented and labelled phonetically and prosodically by two transcribers. The

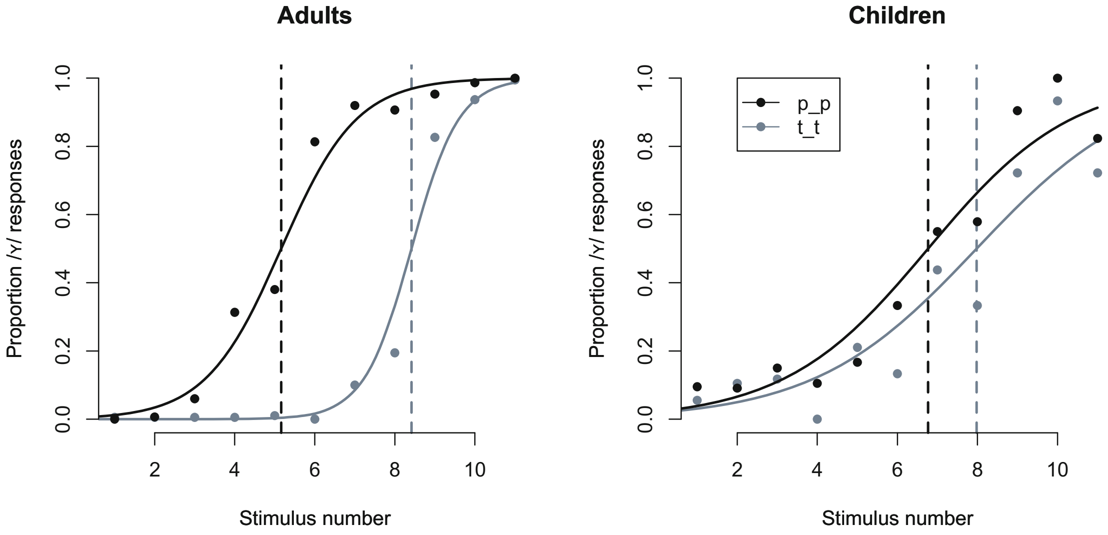

Fig. 4 Psychometric curves showing the proportion of /ʏ/ responses as a function of stimulus number calculated across all subjects for adults (*left*) and children (*right*) in /p_p/ (black) and /t_t/ (*grey*) contexts. The points are proportions at each stimulus number averaged across all subjects (19 adults, 8 children). The *vertical lines* show the 50 % cross-over boundaries at which the proportion of /ʊ, ʏ/ responses are both equal to 0.5. The increasing stimulus numbers extend from low to high F2

acoustic vowel boundaries were marked at the onset and offset of periodicity of each vowel. The formant frequencies were calculated with a 12.5 ms Blackman window and a frame shift of 5 ms. The formant data was checked manually and hand-corrected if necessary. All mis-imitations (e.g. /p/ for /t/ substitutions) and target words that were not phrase-medial were excluded from further analyses (a total of 354 out of 858 tokens).

In order to be able to compare the adult and child data, the child imitations had to be converted into /ʏ/ or /ʊ/ categorical responses. For this purpose, training was carried out on the child-production database and testing involved classifying each imitation as one of these two vowel categories. We included only children who had produced a minimum of 5 /ʊ/ and 5 /ʏ/ vowels, since otherwise it was difficult to achieve statistical stability in constructing the training models. Since three children had produced less than this number of tokens, training and testing were carried out on the data from the remaining 10 children and only the data from those 10 children were further analysed below.

The training and testing were accomplished separately for each child. Training was a Gaussian classification [13, 26, 58] based on the first two formant frequencies at the acoustic temporal midpoint of the vowel. Testing was a maximum likelihood classification based on whichever Bayesian distance to the two vowel categories was smallest.

4 Results

We first removed responses from those combinations of listeners and consonantal-contexts in which there was no convergence in the psychometric curves, i.e. if the resulting decision boundary for any given listener on either continuum fell outside the range of the stimuli (i.e. outside the range of the x-axis shown in Fig. 4). This happened if e.g. a listener responded to a continuum almost entirely with either /ʊ/ or /ʏ/. This required removing all ($n = 6$) response data from 2 children and 1 adult (leaving data from 8 children and 19 adults for further analysis). In addition, responses to the p_p continuum from a further 4 adults and one child ($n = 5$) and to the t_t continuum from additionally 1 child ($n = 1$) had to be removed for the same reason. Thus of the 60 possible original decision boundaries ((20 adults + 10 children) × 2 consonantal contexts), 48 decision boundaries and their associated perceptual responses remained and were analysed below, after removing these data.

As schematically outlined in Fig. 2 above, the greater the distance between the decision boundaries of the psychometric curves, the more listeners normalised for context, i.e. the more they perceived the same acoustic stimulus to be different in the two contexts. The results of the group psychometric curves in Fig. 4 clearly show a greater contextual normalisation for adults than for children. The same figure also shows that the psychometric curves are a good deal flatter for the children which means that they perceived the /ʊ-ʏ/ phonological contrast less distinctively than did adults.

jmh@phonetik.uni-muenchen.de

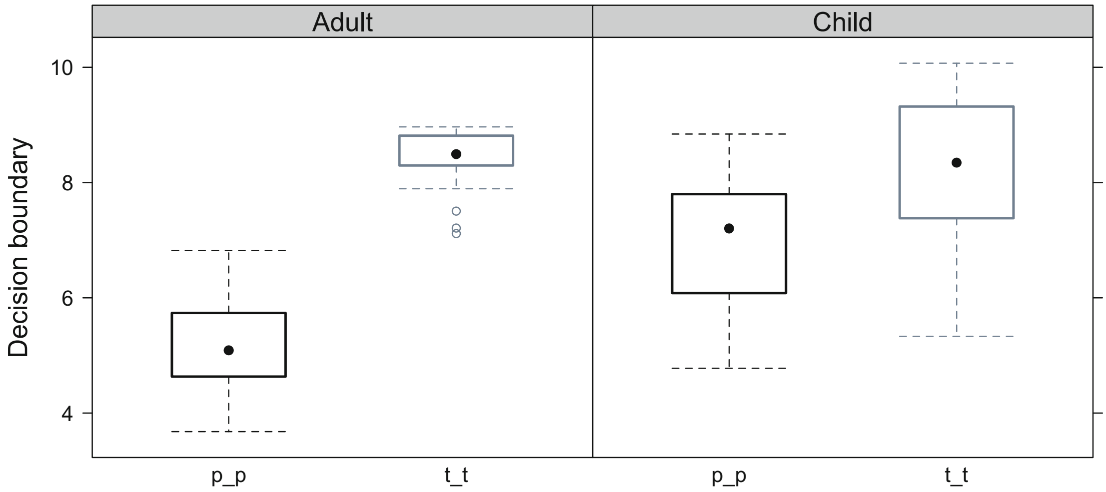

Fig. 5 50 % decision boundaries (corresponding to the *vertical dashed lines* in this figure) for 19 adults (*left*) and 8 children (*right*) in the p_p and t_t contexts. There is one data point per listener in each of the four distributions

The individual decision boundaries for the 19 adult and 8 child listeners in Fig. 5 below show, consistently with the group plots in Fig. 4, a much greater separation between the labial and alveolar contexts for adults than for children. Since the results of a mixed model with the decision boundary as the dependent variable, with consonantal context (2 levels: labial, alveolar) and group (2 levels: adult, child) and with the listener as a random factor showed a significant interaction between the fixed factors ($\chi^2_1 = 9.3$, $p < 0.01$), we applied post-hoc Tukey tests to the same data. These showed consistently with Fig. 5 a significant difference between the decision boundaries for adults ($z = 9.1$, $p < 0.001$) but not for children. They also showed a significant difference in the decision boundaries between adults and children in the labial ($z = 3.7$, $p < 0.001$) but not in the alveolar context. Thus as Fig. 5 shows, the decision boundaries in the labial context are much closer to those in the alveolar context for children than for adults.

5 General Discussion

The study has been concerned with the influence of segmental context on the perception of phonological contrasts and the way in which normalisation (compensation) for context can be affected by variation in speaking style (the first experiment) or by differences between speaker groups (the second). The experiments were conceived within the theory being developed in this chapter that parity between how coarticulation is produced and perceived represents a stable association between phonological categories and speech signals, whereas sound change can come about when there is a misalignment between the modalities in processing coarticulation.

jmh@phonetik.uni-muenchen.de

The first experiment tested whether such instability was more likely in hypoarticulated signals. The sound change concerned a case of dissimilation by which historical /w/ has been deleted when there is another /w/ that occurs later in the same word. Hypoarticulation was simulated by a post-focal, deaccented synthesis of the target word. The results showed fewer initial /w/ perceptions in the hypoarticulated condition. The fewer perceptions of /w/ in the hypoarticulated condition were not due to the presence or absence of a following /w/ (in the next word), but were instead explained as a result of confusion with the lip-rounding that was simulated synthetically by lowering the second formant frequency throughout the target word. Our interpretation of these results is that in hypoarticulated speech, listeners less effectively parsed the signal into those properties that were due to the initial /w/ and those that came about because of the simulated long-range lip-rounding: that is, in a hypoarticulated speaking style phonetic long-range lip-rounding is more likely to mask or camouflage perceptually the initial /w/ than in hyperarticulated speech.

The second experiment built upon earlier findings by Nittrouer and Studdert-Kennedy [47] to test whether the perceptual adjustments for consonantal context are weaker in children than in adults. Our results were consistent with this hypothesis: adults' decisions were swayed to a greater extent by consonantal context than those of children in categorising German lax high rounded vowels.

We now consider the extent to which our interpretation of these results is consistent with Ohala's [50, 51] theory concerning the conditions under which sound change is likely to occur. In our model, dissimilation resulting from a deletion of the initial /w/ comes about because listeners cannot so easily distinguish the initial /w/ from the effects of long-range lip-rounding in a hypoarticulated speaking style. Thus, in contrast to Ohala, we do not invoke an over-normalisation for coarticulation to explain either these data or the processes leading to dissimilation in general. Further, our model differs from Ohala's because we propose that long-range lip-rounding itself rather than the presence of a second /w/ is sufficient to trigger the perceptual deletion of the first /w/. On the other hand, our interpretation that dissimilation comes about as a result of long-range lip-rounding perceptually masking or hiding a segment that shares the same phonetic properties is very reminiscent of Ohala's [48] view that there is commonality between the mechanisms leading to dissimilation and those in the visual domain causing an object to be camouflaged (as when—to use his analogy—a white rabbit is hidden against a background of snow). The difference here is that our model incorporates speaking style: the perceptual masking that can lead to dissimilation is more likely to happen in response to hypoarticulated speech.

The data from the second experiment are relevant to Ohala's [50] interpretation that many more sound changes arise because of an *insufficient* perceptual normalisation for coarticulation. With regard to the data in the second experiment, Ohala's [50] idea is that the back vowel of /tut/ can change category for the listener into a front vowel /tyt/ when the listener no longer normalises for coarticulation: that is, the prior stage in Ohala's model to diachronic category change is that /u/ in /tut/ is decontexualised such that a listener ceases to apply a perceptual shift to compensate for the phonetic raising effects due to the flanking alveolars. With regard to the model schematised in Fig. 2, Ohala's prediction of perceptual under-compensation for coarticulation is that the right /tʊt-tʏt/ boundary should shift to the left so that the section marked ‘normalise’ is recategorised as /ʏ/. But this is not what our data show. Recall that there was no difference between children and adults in the location of the /tʊt-tʏt/ boundary: children’s diminished normalisation for context instead came about because of a rightwards shift of the /pʊp-pʏp/ boundary towards that of /tʊt-tʏt/.

So do these data from this second experiment have anything to say about sound change? We think that they do. This is because they form a consistent pattern with our other various different types of analyses of /ʊ, u/-fronting based on (i) longitudinal studies in the same individual [25], (ii) an apparent-time study comparing younger and older speakers and listeners [27] in Standard Southern British, and more recently (iii) normalisation for /ʊ/-fronting in deaccented speech in German [28]. Whenever our perception analyses showed evidence for an under-compensation for coarticulation, then, just as for the children in the present experiment, the decision boundary of the /ʊ/ or /u/-variants in the back or non-fronting context shifted to the front—and not as predicted by Ohala’s [50] model the other way round. Our findings from perception in these studies are also generally consistent with production data (e.g. for the longitudinal study in Harrington [25]) by which there was a shift of non-fronting (e.g. ‘m**o**ve’, /muv/) towards fronted /u/-variants (e.g. ‘m**u**te’, /mjut/).

Our explanation for these findings across these studies is that diachronic /ʊ, u/ fronting is the outcome of a synchronically gradual shift induced by /ʊ, u/ undershoot which pushes non-fronting (‘m**o**ve’) towards fronted (‘m**u**te’) variants. Acoustically, the shift is brought about by an F2-raising of non-fronting /ʊ, u/-variants due to their centralisation under hypoarticulation. That is, acoustically the F2-space of vowels in words like ‘move’ becomes more extensive when there is target undershoot/hypoarticulation: crucially, this extension due to hypoarticulation/undershoot is *asymmetric* towards higher F2 values. The perceptual response to this greater asymmetric variation in production is firstly the flatter psychometric response curve observed for the labial context in the present study for adults compared with alveolars (compare the steepness of the two sigmoids in the left panel of Fig. 4). Secondly, if inexperienced listeners are predominantly exposed in speech communication to hypoarticulated back vowels, then it is not just the sigmoid slope of the perceptual contrast that will decrease: the decision boundary will also shift up the F2 scale towards the fronted variant. This we would suggest is the reason why children’s perceptual responses in the labial context are shifted much further to the right and nearer to those of the alveolar context than for adults. Based on these data then, diachronic /ʊ, u/-fronting is not brought about as Ohala [50] has argued because a listener gives up compensating for coarticulation (which should cause the alveolar decision boundary to shift towards the labial one). It is instead the outcome of phonologising a back vowel variant (‘m**o**ve’) that extends synchronically due to hypoarticulation towards a variant (‘m**u**te’) that is already front as a result of coarticulation.

The core idea here is then that hypoarticulated variants shift towards those that are substantially affected by coarticulation. To what extent does this idea generalise to other types of sound change? For Ohala [50], sound changes such as diachronic /u/-fronting, umlaut resulting from VCV coarticulation, and the development of vowel nasalisation are all brought about if a listener turned speaker normalises insufficiently for coarticulation. In our model, all of these sound changes are linked not by insufficient compensation for coarticulation, but by the weakening effects induced by hypoarticulation. In the absence of data of our own, we speculate that hypoarticulation can just as easily target the source of coarticulation as the coarticulatory effect, causing its weakening (e.g. modern Standard German *Füße* with a final /ə/ derived from Old High German /fotiz/) or its deletion (as in French 'main', /mɛ̃/ from Latin 'manus'). This weakening/deletion of the source would lead to its decoupling from (and eventual phonologisation of) the coarticulatory effect. Thus these three sound changes share in common that they are all derived not from a perceptual under-compensation for coarticulation but instead from the perceptual consequences of hypoarticulated speech signals: the main difference across these sound changes is that hypoarticulation in /u/-fronting targets the sound that is ultimately changed, whereas in VCV coarticulation and the phonologisation of vowel nasalisation, it targets the source that gives rise to the sound change.

Hypoarticulation is also the synchronic factor that links the two experimental findings of this chapter: in the first, a hypoarticulated speaking style causes a perceptual camouflage between long-range anticipatory coarticulation and an initial consonant with similar phonetic properties, potentially leading diachronically to its dissimilation. In the second, hypoarticulation causes synchronically a bias that shifts back /ʊ/-variants towards their fronted variants, leading potentially to their merger and diachronic recategorisation as front vowels. This merger was evident in the children's responses in Experiment 2. The link between both experiments is the idea that hypoarticulation results in listener uncertainty. This is substantiated by the evidence showing flatter sigmoids for deaccented versus accented responses for adults (Experiment 1: Fig. 1) and for children versus adults (Experiment 2: Fig. 4).

In our proposed model, sound change arises synchronically out of the interaction of the separate forces of hypoarticulation and coarticulation that act upon the transmission of speech. From this point of view, our model integrates the insights expressed respectively in Lindblom et al. [44] and Ohala [50, 51] that both speaking style variation and the perception of coarticulation together set the conditions for sound change to occur. Our model is consistent with past [8, 9, 53, 55] as well as more recent [31, 41, 65] findings that, because lexically more frequent words tend to be hypoarticulated relative to less frequent ones [18, 42, 62], then sound change may often be lexically gradual and dependent on lexical statistics.

Finally, our model in which hypoarticulation plays a central role in setting the conditions for sound change to occur is also consistent with the evidence that sound change is very often reductive. But we emphasise that our model does not predict that sound change *must* be reductive. The data from the second experiment represent just such an example in which diachronic vowel recategorisation emerges out of hypoarticulation pushing (in both perception and production) back towards coarticulatory fronted high vowel variants without any vowel reduction.

Acknowledgments This research was supported by European Research Council grant number 295573 'Sound change and the acquisition of speech' (2012–2017).

References

1. Abrego-Collier, C.: Liquid dissimilation as listener hypocorrection. In: Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society, pp. 3–17 (2013)
2. Alderete, J., Frisch, S.: Dissimilation in grammar and the lexicon. In: de Lacy, P. (ed.) The Cambridge Handbook of Phonology. Cambridge University Press, Cambridge (2006)
3. Beckman, M., Li, F., Kong, E., Edwards, J.: Aligning the timelines of phonological acquisition and change. Lab. Phonol. **5**, 151–194 (2014)
4. Beddor, P.: A coarticulatory path to sound change. Language **85**, 785–821 (2009)
5. Beddor, P.: Perception grammars and sound change. In: Solé, M.-J., Recasens, D. (eds.) The Initiation of Sound Change. Perception, Production, and Social Factors, pp. 37–55. John Benjamin, Amsterdam (2012)
6. Blevins, J.: Evolutionary Phonology: The Emergence of Sound Patterns. Cambridge University Press, Cambridge (2004)
7. Browman, C., Goldstein, L.: Articulatory phonology: an overview. Phonetica **49**, 155–180 (1992)
8. Bybee, J.: Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Lang. Var. Change **14**, 261–290 (2002)
9. Chen, M., Wang, W.: Sound change: actuation and implementation. Language **51**, 255–281 (1975)
10. Cutler, A., Foss, R.: On the role of sentence stress in sentence processing. Lang. Speech **20**, 1–10 (1977)
11. de Jong, K.: The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. J. Acoust. Soc. Am. **97**, 491–504 (1995)
12. D'Imperio, M.: Italian intonation: an overview and some questions. Probus **14**(1), 37–69 (2002)
13. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. Wiley, New York (2001)
14. Foulkes, P., Vihman, M.: Language acquisition and phonological change. In: Honeybone, P., Salmons, J.C. (eds.) The Handbook of Historical Phonology. OUP, Oxford (in press)
15. Fowler, C.: Segmentation of coarticulated speech in perception. Percept. Psychophys. **36**, 359–368 (1984)
16. Fowler, C.: Speech as a supramodal or amodal phenomenon. In Calvert, G., Spence, C., Stein, B. (eds.) The Handbook of Multisensory Processes, pp. 189–201. MIT Press, Cambridge (2004)
17. Fowler, C.: Parsing coarticulated speech in perception: effects of coarticulation resistance. J. Phonetics **33**, 199–213 (2005)
18. Fowler, C., Housum, J.: Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. J. Memory Lang. **26**, 489–504 (1987)
19. Fowler, C., Smith, M.: Speech perception as "vector analysis": an approach to the problems of segmentation and invariance. In: Perkell, J., Klatt, D. (eds.) Invariance and Variability in Speech Processes, pp. 123–139. Erlbaum, Hillsdale (1986)
20. Fowler, C., Thompson, J.M.: Listeners' perception of "compensatory shortening". Attention Percep. Psychophys. **72**, 481–491 (2010)
21. Fujisaki, H., Kunisaki, O.: Analysis, recognition and perception of voiceless fricative consonants in Japanese. Annu. Bull. Res. Inst. Logop. Phoniatr. **10**, 145–156 (1976)
22. Grammont, M.: Observations sur le langage des enfants. In: Mélanges Linguistiques Offerts aI M. Antoine Meillet, pp. 115–131. Klincksieck, Paris (1902)
23. Grosvald, M.: Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. J. Phonetics **37**, 173–188 (2009)
24. Grosvald, M., Corina, D.: The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In: Solé, M.-J., Recasens, D. (eds.) The Initiation of Sound Change. Perception, Production, and Social Factors, pp. 77–100. John Benjamins, Amsterdam (2012)

jmh@phonetik.uni-muenchen.de

25. Harrington, J.: Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In: Cole, J., Hualde, J. (eds.) Laboratory Phonology, vol. 9, pp. 125–143. Mouton, Berlin (2007)
26. Harrington, J.: The Phonetic Analysis of Speech Corpora. Wiley-Blackwell (2010)
27. Harrington, J., Kleber, F., Reubold, U.: Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study. J. Acoust. Soc. Am. **123**, 2825–2835 (2008)
28. Harrington, J., Kleber, F., Reubold, U.: The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. J. Acoust. Soc. Am. **134**, 551–561 (2013)
29. Harrington, J., Kleber, F., Reubold, U., Siddins, J.: The implications for sound change of prosodic weakening: evidence from polysyllabic shortening. J. Lab. Phonol. **6**, 81–117 (2015)
30. Hay, J., Jannedy, S., Mendoza-Denton, N.: Oprah and /ay/: lexical frequency, referee design and style. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco (1999)
31. Hay, J., Foulkes, P.: The evolution of medial /t/ over real and remembered time. Language (in press)
32. Heid, S., Hawkins, S.:. An acoustical study of long domain /r/ and /l/ coarticulation. In: Proceedings of the 5th Seminar on Speech Production: Models and Data, pp. 77–80. Kloster Seeon, Bavaria, Germany, Munich (2000)
33. Kataoka, R.: Phonetic and cognitive bases of sound change. Ph.D. dissertation, University of California, Berkeley (2011)
34. Kawasaki, H.: Phonetic explanation for phonological universals: the case of distinctive vowel nasalization. In: Ohala, J., Jaeger, J. (eds.) Experimental Phonology, pp. 81–103. Academic Press, Orlando (1986)
35. Kiparsky, Paul: Universals constrain change; Change results in typological generalizations. In: Good, Jeff (ed.) Linguistic Universals and Language Change, pp. 23–53. Oxford University Press, Oxford (2008)
36. Kleber, F., Harrington, J., Reubold, U.: The relationship between the perception and production of coarticulation during a sound change in progress. Lang. Speech **55**, 383–405 (2012)
37. Kleber, F., Peters, S.: Children's imitation of coarticulatory patterns in different prosodic contexts. In: Proceedings of 14th Conference on Laboratory Phonology, Tokyo, Japan (2014)
38. Labov, W.: Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia (1972)
39. Ladd, D.: Intonational Phonology, 2nd edn. Cambridge University Press, Cambridge (2008)
40. Lightfoot, D.: The Development of Language: Acquisition, Change, and Evolution. Blackwell, Malden (1999)
41. Lin, S., Beddor, P., Coetzee, A.: Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. Lab. Phonol. **5**, 9–36 (2014)
42. Lindblom, B.: Explaining phonetic variation: a sketch of the H & H theory. In: Hardcastle, W., Marchal, A. (eds.) Speech Production and Speech Modeling, pp. 403–439. Kluwer, Dordrecht (1990)
43. Lindblom, B.: Systemic constraints and adaptive change in the formation of sound structure. In: Hurford, J., Studdert-Kennedy, M., Knight, C. (eds.) Approaches to the Evolution of Language: Social and Cognitive Bases, pp. 242–264. Cambridge University Press, Cambridge (1998)
44. Lindblom, B., Guion, S., Hura, S., Moon, S.-J., Willerman, R.: Is sound change adaptive? Rivista di Linguistica **7**, 5–36 (1995)
45. Lindblom B., MacNeilage, P.: Coarticulation: A universal phonetic phenomenon with roots in deep time. FONETIK 2011, 41–44 in TMH—QPSR, vol. 51, KTH Stockholm (2011)
46. Mann, V., Repp, B.: Influence of vocalic context on the perception of [ʃ]-[s] distinction: I. Temporal factors. Percep. Psychophys. **28**, 213–228 (1980)
47. Nittrouer, S., Studdert-Kennedy, M.: The role of coarticulatory effects in the perception of fricatives by children and adults. J. Speech Hear. Res. **30**, 319–329 (1987)
48. Ohala, J.: The listener as a source of sound change. In: Masek, C.S., Hendrick, R.A., Miller, M.F. (eds.) Papers from the Parasession on Language and Behavior, pp. 178–203. Chicago Linguistic Society, Chicago (1981)

49. Ohala, J.: The phonetics of sound change. In: Jones, C. (ed.) Historical Linguistics: Problems and Perspectives, pp. 237–278. Longman, London (1993)
50. Ohala: The listener as a source of sound change: an update. In: Solé, M.-J., Recasens, D. (eds.) The Initiation of Sound Change. Perception, Production, and Social factors, pp. 21–36. John Benjamins, Amsterdam (2012)
51. Ohala, J., Feder, D.: Listeners' identification of speech sounds is influenced by adjacent "restored" phonemes. Phonetica **51**, 111–118 (1994)
52. Öhman, S.E.: Coarticulation in VCV utterances: spectrographic measurements. J. Acoust. Soc. Am. **39**, 151–168 (1966)
53. Phillips, B.: Word Frequency and Lexical Diffusion. Palgrave Macmillan, Basingstoke (2006)
54. Pierrehumbert, J., Hirschberg, J.: The meaning of intonational contours in the interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M. (eds.) Intentions in Communication, pp. 271–311. MIT Press, Cambridge (1990)
55. Pierrehumbert, J.: Exemplar dynamics: word frequency, lenition, and contrast. In: Bybee, J., Hopper, P. (eds.) Frequency Effects and the Emergence of Lexical Structure, pp. 137–157. John Benjamins, Amsterdam (2001)
56. Plichta, B.: AKUSTYK for Praat [Computer program] (2010). Accessed 12 June 2015
57. Solé, M.: The perception of voice-initiating gestures. Lab. Phonol. **5**, 37–68 (2014)
58. Srivastava, S., Gupta, M., Frigyik, B.: Bayesian quadratic discriminant analysis. J. Mach. Learn. Res. **8**, 1, 277–305 (2007)
59. Studdert-Kennedy, M.: Introduction: the emergence of phonology. In: Hurford, J., Studdert-Kennedy, M., Knight, C. (eds.) Approaches to the Evolution of Language, pp. 169–176. Cambridge University Press, Cambridge
60. Vihman, M.: Sound change and child language. In: Traugott, E., La Brum, R., Shepherd, S. (eds.) Papers from the 4th International Conference on Historical Linguistics, pp. 303–320. John Benjamins, Amsterdam (1980)
61. West, P.: Perception of distributed coarticulatory properties of English /l/ and /r/. J. Phonetics **27**, 405–425 (2000)
62. Wright, R.: Factors of lexical competition in vowel articulation. In: Local, J., Ogden, R., Temple, R. (eds.) Papers in Laboratory Phonology VI, pp. 75–87. Cambridge University Press, Cambridge (2003)
63. Yu, A.: Individual differences in socio-cognitive processing and the actuation of sound change. In: Yu, A. (ed.) Origins of Sound Change: Approaches to Phonologization, pp. 201–227. Oxford University Press, Oxford (2013)
64. Yu, A., Abrego-Collier, C., Sonderegger, M.: Phonetic imitation from an individual-difference perspective: subjective attitude, personality, and 'autistic' traits. PLoS One **8**(9), e74746 (2013)
65. Zellou, G., Tamminga, M.: Nasal coarticulation changes over time in Philadelphia English. J. Phonetics **47**, 18–35 (2014)

jmh@phonetik.uni-muenchen.de