In Anna Esposito & Lakhmi Jain (Editors) Toward Robotic Socially Believable Behaving Systems: Modeling Social Signals. ISBN 978-3-319-31052-7. Springer. (p. 61-85).

Chapter 6
The Relevance of Context and Experience for the Operation of Historical Sound Change

Jonathan Harrington, Felicitas Kleber, Ulrich Reubold and Mary Stevens

Abstract This paper is concerned with explaining how historical sound change can emerge as a consequence of the association between continuous, dynamic speech signals and phonological categories. The relevance of this research to developing socially believable speech processing machines is that sound change is both cognitive and social and also because it provides a unique insight into how the categories of speech and language and dynamic speech signals are inter-connected. A challenge is to understand how unstable conditions that can lead to sound change are connected with the more typical stable conditions in which sound change is minimal. In many phonetic models of sound change, stability and instability come about because listeners typically parse—very occasionally misparse—overlapping articulatory movements in a way that is consistent with their production. Experience-based models give greater emphasis to how interacting individuals can bring about sound change at the population level. Stability in these models is achieved through reinforcing in speech production the centroid of a probability distribution of perceived episodes that give rise to a phonological category; instability and change can be brought about under various conditions that cause different category distributions to shift incrementally and to come into contact with each other. Beyond these issues, the natural tendency to imitation in speech communication may further incrementally contribute to sound change both over adults' lifespan and in the blending of sounds that can arise through dialect contact. The general conclusion is that the instabilities that give rise to sound change are an inevitable consequence of the same mechanisms

J. Harrington (✉) · F. Kleber · U. Reubold · 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

U. Reubold
e-mail: mes@phonetik.uni-muenchen.de

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

A. Esposito and L.C. Jain (eds.), *Toward Robotic Socially Believable Behaving Systems - Volume II*, Intelligent Systems Reference Library 106,
DOI 10.1007/978-3-319-31053-4_6

that are deployed in maintaining the stability between phonological categories and their association with the speech signal.

A fundamental challenge in phonetics and the speech sciences is to understand how the interleaved movements of continuous speech signals are associated with categories such as consonants and vowels that function to distinguish between word meanings. The dichotomy between these two levels of representation comes about because on the one hand any speech utterance is highly context-dependent but on the other hand languages distinguish words by means of a finite cipher of abstract phonological units that can be permuted in different ways. There is abundant evidence for the context-dependent nature of speech. The same utterance can vary dramatically depending on the speaking situation and environment—whether talking to friends or in a more formal speaking situation, whether there is background noise or quiet [92]. Speech is also context-dependent because speech sounds are synchronised in a temporally overlapping way: producing speech is a shingled movement [144], so that any particular time slice of the speech signal provides the listener with information about speech sounds that have been, and that are about to be produced and in a way that is also different depending on prosodic factors to do with syllable position and the stress or prominence with which syllables are produced [9]. These are then many of the reasons why temporally reversing a speech signal of *stack* does not lead to an unambiguous percept of *cats*. Moreover, the context-dependence is not just a function of the sluggishness of the vocal organs in relation to the speed with which speech is produced, but also communicates much about the social and regional affiliations of the speaker [60, 96]. At the same time, phonological abstraction from these details is fundamental to human speech processing: there is a sense in which the different words *stack*, *cats*, *acts*, and *scat* are permutations of the same four phonological units or phonemes. There is now extensive evidence that children learn both levels of representation in speech communication: they are on the one hand responsive to acoustic information arising from continuous movement in the speech signal. But simultaneously they acquire the ability to perform phonological abstraction which allows them to recombine abstract phonological units to produce words that they have not yet encountered (e.g. Beckman et al. [10]).

The task in this paper is to make use of the existing knowledge about the connections between these two very different ways of representing speech in order to explain the operation of sound change; and in turn to use what is known about sound change to place constraints on the type of architecture that is possible and required for linking these physical and abstract levels of speech sound representation. The focus will be on what the Neogrammarians of the 19th century [115, 118] termed regular sound change which they considered to be gradual and imperceptible and to apply to all words; this type of sound change was for them distinct from analogical change which was irregular, phonetically abrupt (in the sense that the change was immediate, not gradual), lacked phonetic motivation, and often applied to only a handful of words (see e.g. Hualde [68] for some examples of change by analogy).

Modelling sound change is relevant to understanding how cognitive and social aspects of human speech processing are connected. The association between these two domains has been largely neglected in the 20th century, partly because whereas generative theories of phonology draw upon highly idealised data (typically phonetic transcriptions) to develop a grammar consisting of re-write rules based on supposed linguistic universals and operating on a lexicon containing the minimum of information to represent the distinctions between words, sociolinguistic models by definition are concerned with explaining how variation in speech is conditioned by factors such as gender, age, and social class, factors that are beyond the scope of generative models. In the last 10–15 years, there have, however, been increasing attempts to reconcile these two positions largely within so-called usage-based, episodic or exemplar models of speech (see e.g. Docherty and Foulkes [31] for a recent review) that are derived from models of perception in cognitive psychology (e.g. Hintzman [65]) and that give much greater emphasis to the role of memory in human speech processing. Exemplar theory has led to computational models of how phonological categories, the lexicon, memory and speech are inter-connected (e.g. Wedel [157]); and more generally, there has been greater emphasis in the last two decades in determining how speaker-specific attributes shape and influence cognitive aspects of human speech processing both in adults [122] and in first language acquisition (e.g. Beckman et al. [10], Munson et al. [103]). Understanding how social and cognitive aspects are related in human speech processing is in turn a pre-requisite for developing socially believable systems of machine speech and language processing.

The relevance of sound change in this regard is that it is evidently both cognitive and social. The cognitive aspects are largely concerned with the mechanisms by which phonological categories and speech signals are associated and how this association can sometimes become unstable providing the conditions for sound change to take place. The social aspects are more concerned with how differences between speakers in their knowledge and use of language can cause sound change to spread throughout the community. These cognitive and social bases of sound change have in the last 40–50 years been pursued within largely separate frameworks concerned on the one hand with the conditions that can give rise to sound change (in particular Ohala [112]) and those that lead to its spread across different speakers on the other (in particular Labov [85, 87]). The challenge lies in developing an integrated model of sound change that draws upon insights from both approaches. This is turn can provide fresh insights into understanding how social and cognitive aspects must be inter-connected in both human and (therefore also) machine speech processing.

6.1 The Phonetic Basis of Sound Change

Much consideration has been given to the question of whether there are factors intrinsic to the structure of the language that can bring about sound change. Such questions are typically focussed on whether there are sounds and in particular sequences of sounds that are inherently unstable, either for reasons to do with speech production or because they tend to be ineffectively communicated to the listener. Such biases in either the production or the perception of speech that predispose sounds to change should also be reflected in the typological distribution of sounds and sound sequences in the languages of the world. Thus syllables beginning with /kn, gn/ as in German *Knabe*, *Gnade* are rarer in the languages of the world than those beginning with /kl, gl/ [67]; and they are also involved in sound changes by which /k, g/ are deleted as in the evolution of English words that have fossilized the earlier (16th century) pronunciations in the orthography (*knave*, *knight, knife*, *gnome*, *gnat etc.*) but which are now in all English varieties pronounced without the initial velar stop [97].

Research concerned with the structural conditions that give rise to sound change is founded on two further ideas. Firstly, the biases that bring about sound change are directional [113]. Thus there are many sound changes by which the front vowel /y/ has developed historically from a back vowel /o/ or /u/ (e.g. the modern German *Füße*, with /fys/ in the first syllable historically from Proto-Germanic /fotiz/; hence also the English alternation *feet*, *foot*) but far fewer sound changes by which /y/ has retracted with the passage of time to /u/ or /o/: that is, there is a bias towards high back vowel fronting (as opposed to high front vowel retraction) both synchronically and diachronically [55]. Similarly, Guion [48] has shown that the misidentification of /ki/ as /tʃi/ which forms the basis of sound changes known as velar palatalization in numerous languages (e.g. English *chin*, but German *Kinn*) is far more likely than the perhaps unattested sound change by which /tʃi/ evolves into /ki/. Secondly, to the extent that speakers from different languages and cultures are endowed with the same physical mechanisms for producing and perceiving speech, there should be broadly similar patterns of sound change (sometimes referred to as regular sound change) in unrelated languages.

The model of sound change developed by Ohala [111–113] over a number of decades is founded upon such principles. The basis for this model is that coarticulation in speech production—that is the way in which speech sounds overlap and influence each other in time—is in almost all cases accurately transmitted between speakers and hearers. An example of coarticulation in speech production is given in Fig. 6.1 which shows the distribution of the back vowel /ʊ/ (e.g. *m<u>u</u>sste*, 'had to') in non-fronting /p/ and fronting /t/ contexts in German. In the fronting context, the tongue dorsum for /ʊ/ is further forward in the mouth due to the influence of /t/ that has its primary constriction further forward than that of /ʊ/ resulting acoustically in a raised second formant (F2) frequency.

How does the listener deal with the type of variation shown in Fig. 6.1? Experiments over several decades [41, 93, 99] show that adult listeners of the language interpret speech production in relation to the context in which it was produced. For the present example, this implies that listeners carry out a type of perceptual transformation such that they attribute the coarticulatory fronting not to the vowel itself but to the consonantal context in which it was produced. This can be demonstrated in perception experiments by synthesising a continuum between a front and back vowel in equal steps and embedding the continuum in coarticulatory fronting and non-fronting contexts. Figure 6.2 shows the results from just such an experiment for the fronting context /jist-just/, *yeast-used* (past tense) and for the non-fronting context /swip-swup/, *sweep-swoop*. (The former is a fronting context because of the /j/ and the latter a non-fronting context because the tongue dorsum for /w/ is retracted,

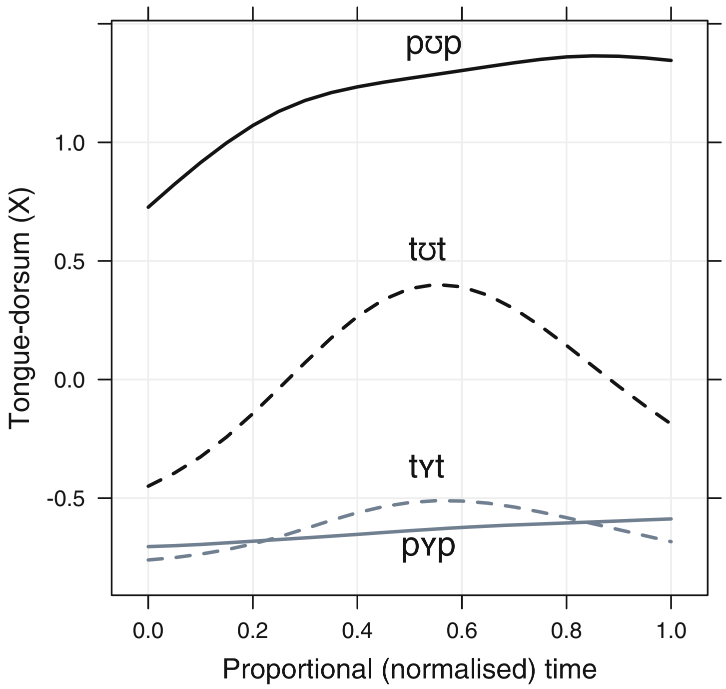

Fig. 6.1 The horizontal position of the tongue dorsum as a function of normalized time extending between the acoustic onset and offset of the /ʊ, Y/ (*black*, *grey*) in /p, t/ (*solid*, *dashed*) contexts in German non-words. The data are taken from Harrington et al. [56] and were aggregated across seven speakers of German following speaker normalization using z-score [95] normalization. Higher/lower (increasingly positive/negative) values on the *y*-axis are increasing back/front positions respectively of the tongue dorsum

as for /ʊ/). The results in Fig. 6.2 show the responses from listeners who carried out a forced choice test i.e. identified each stimulus as one of the words. As Fig. 6.2 shows, listeners were more inclined to perceive /u/ in the fronting (*yeast-used*) context. Presumably, this is because they attributed some of the coarticulatory fronting to the consonantal context itself and so biased their responses towards /u/. Another way of putting this is to say that listeners factored out from the acoustic signal the part that was caused by coarticulatory fronting and associated or *parsed* it with the source that gives rise to it, the consonantal context. More generally, listeners of speech have to associate or parse the acoustic consequences of the speaker's interleaved or 'shingled' movements: for the example in Fig. 6.2, they are interleaved because the action of tongue fronting due to the consonantal context is produced or overlaid on the motor actions that are required for the production of the vowel. Models such as articulatory phonology [21] and its forerunner action theory [36, 39] are founded on the premise of a parity between the production and perception modalities: that is, listeners' parsing of the coarticulatory information (/u/-fronting) with the source that is responsible for it (an anterior consonant) is a consequence of their direct perception of the interleaved or shingled movements produced by the speaker.

Ohala's [111, 112] insight is that occasionally the listener does not parse the speech signal in relation to phonological units consistently with their production. A well-known example is coarticulatory nasalization in VN (vowel + nasal

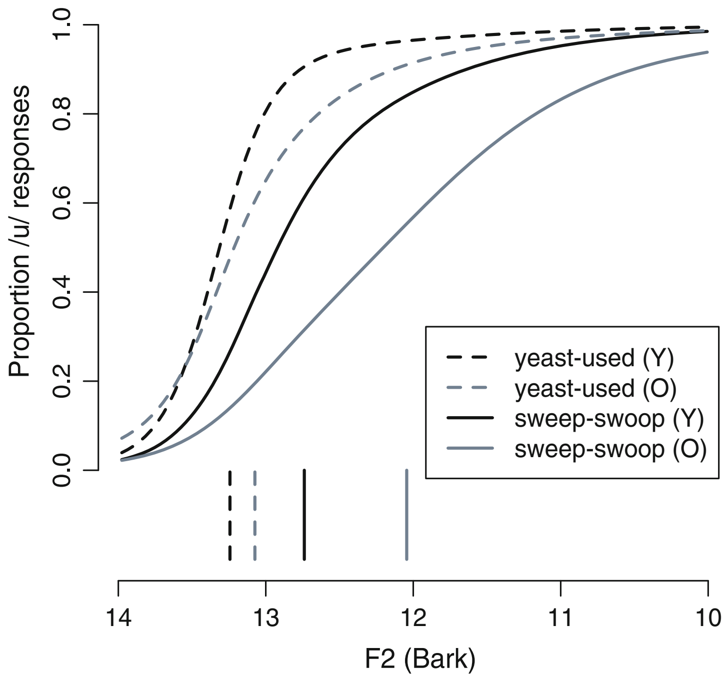

Fig. 6.2 Psychometric curves showing decisions of younger (*black*) and older (*grey*) listeners to continua synthesised between /i, u/ and embedded in fronting (*yeast-used*) and non-fronting (*sweep-swoop*) contexts. The vertical axis shows the proportion of responses identified as /u/ as opposed to /i/. The horizontal axis extends between acoustically most /i/-like (*left*) to acoustically most /u/-like (*right*) vowels. The *vertical lines* mark the cross-over boundaries i.e. where responses were equivocal between /i, u/. Adapted from Harrington et al. [56]

consonant) sequences. If speakers and hearers agree on how the signal is to be parsed into phonological units, then a vowel that is nasalised because it is produced in the context of nasal consonants should be perceived to be oral. Just this has been demonstrated experimentally (e.g. Kawasaki [75]). The interpretation of such findings is as follows: listeners perceive a nasalised vowel in a word like *ban* to be just as oral as in *bad* because they parse (or factor out) the contextual nasalisation in the former and attribute it to the source /n/ (see also Beddor and Krakow [14]).

Such would be the interpretation on the assumption of parity between speech production and perception in the processing of coarticulation. However, a set of perception experiments by Beddor [12] shows that listeners are not always consistent in parsing the coarticulation with the source. If they parse only some of the nasalisation with the following N in VN sequences, then they would hear the vowel to be at least partially nasalised. According to Ohala [112], the historical development of vowel nasalisation in languages such as French (e.g. *main*, 'hand', /mɛ̃/ from Latin *manus*) can arise from just this kind of parsing failure that causes some of the nasalisation to be stuck with the vowel (causing it to be perceived as nasal).

Sound changes like the evolution of French /mɛ̃/ from a VN sequence in Latin is in Ohala's model one of *hypo*correction because the listener has not parsed enough of the coarticulatory variation with the source. A sound change such as the insertion of intrusive /p/ in surnames like *Thompson* derived originally from the *son of Thom* (cf. also *glimpse* from Old English *glimsian*) comes about in terms of this model because the acoustic signal in such nasal clusters often contains a silent interval that listeners incorrectly identify as an oral stop. The silent interval between a nasal consonant /m/ and the fricative /s/ is produced for aerodynamic reasons. An /m/ has an oral movement of lip-closure synchronised with a lowered soft palate in which air exits the nasal cavity. But if the soft palate is raised before the closed lips are released, then the /m/ will change into an oral stop, i.e. a /p/ or /b/, which is marked acoustically by a silent interval (with no air exiting the nasal cavity because the soft palate is raised and none exiting the oral cavity because of the lip-closure). The early raising of the soft palate may come about in order to build up sufficient air pressure in the mouth cavity that is in turn required for the production of the following fricative /s/ with a turbulent airstream. Whatever the reason, a silent interval often occurs between nasals and fricatives: this is why English *mince*/*mints*, /mɪm(t)s, mɪmts/ and German *Gans*/*ganz* ('goose'/'quite') /gan(t)s, gants/ are homophonous for most speakers. From the listener's point of view, the permanent sound change is due in Ohala's model to a parsing failure. There could have been no /p/ in the original production of *Thompson* (being derived from the son of *Thom*) but listeners nevertheless insert one because they cannot parse the silent interval with the /m/ or /s/ as neither of these is typically produced with an acoustically silent interval.

Sound change in Ohala's model can also come about due to *hyper*correction in which the listener parses too much coarticulation with the source. Hypercorrection is presumed to be involved in sound changes such as Grassman's law by which aspiration in ancient Greek came to be deleted if the word contained another aspirated consonant (e.g. /**t^{h}**riks/, 'hair' but /**t**rik^{h}os/ 'hair' gen. sing.). In Ohala's model, this sound change of dissimilation (see e.g. Alderete and Frisch [2], Blevins [15], Müller [104]) comes about because listeners incorrectly parse aspiration in the /t^{h}/ of what was presumably an earlier form /t^{h}rik^{h}os/ with the following /k^{h}/ and so factor it out, in much the same way that listeners typically *correctly* factor out nasalization from the vowel in *ban* and parse it with the /n/ (leading to the perception of an oral vowel in *ban*). Although this theory of perceptual dissimilation is plausible and testable, it has so far been difficult to substantiate it by recreating the conditions under which it could have taken place in the laboratory [1, 58, 104]. Harrington et al. [58] provided some evidence that dissimilation sound changes might be explained by the interaction between coarticulation and speaking style in perception. They showed that long-range coarticulatory lip-rounding can in the perception of a hypoarticulated speaking style (see below) mask the perception of a consonant like /w/ that is inherently lip-rounded thereby possibly recreating the synchronic conditions for its diachronic deletion.

Finally, certain types of metathesis in common with dissimilation can occur when the temporal influence of a sound is extensive. In metathesis, two sounds swap their serial position as in modern English *bird* from Old English *bridde* (see Blevins and Garrett [16] and most recently Egurtzegi [32] for copious other examples). Metathesis often involves liquids (i.e. /l, r/) whose acoustic effects are known to have a long time window [64] i.e. to extend often at least throughout the word, perhaps thereby making the identification of their serial position in relation to the other consonants and vowels of the word difficult for the listener. Just this argument has recently been used in

Ruch and Harrington [128] in modelling the change by which post-aspiration has recently developed from pre-aspiration in the Andalusian variety of Spanish (see also Torreira [147]). In Andalusian Spanish, syllable-final /s/ in words like *pasta* is produced not as an /s/, but is debuccalised and has a quality similar to /h/ in some languages, thus [pa^{h}ta]. Older speakers tend to produce [pa^{h}ta] whereas younger speakers are far more inclined to produce [pat^{h}a] in which the aspiration follows the /t/ (as it does in English *ten*). Ruch and Harrington [128] suggest that this sound change in progress may come about because, while listeners are focussed on whether or not aspiration has occurred (in order to distinguish *pasta* 'pasta' from *pata*, 'paw'), they are unable to determine its serial location in relation to the /t/.

There have been several developments to Ohala's model in recent years including in particular the following.

i. *Small articulatory changes can have large acoustic consequences*. This is the basis of the quantal theory of speech production [140, 141] according to which the relationship between speech production and the acoustic signal (and therefore perception) is non-linear. For example, the incremental retraction of the tongue tip towards the hard palate starting from an /s/ is not accompanied by a similarly gradual acoustic change: instead there is a quantal jump (of spectral centre of gravity lowering of the fricative noise in this case) and an abrupt switch in perception from /s/ to /ʃ/. In a recent ultrasound study of the conditions under which /l/ vocalises—resulting synchronically in productions such as London Cockney /waʊ/ for *wall* and diachronically in /l/-deletion (e.g. in *folk*, *palm*, *talk* etc.)—Lin et al. [91] observe that an incremental shift of tongue tip lowering can cause quite a marked acoustic change (of the formant frequencies). Incremental variation in speech production is of course typical (speakers never produce the same utterance exactly identically on two occasions). Perhaps then it is the type of incremental variation that can make a quantal acoustic and perceptual difference which is more likely to evolve into sound change.
ii. *Implementational features* according to Solé [138] facilitate a contrast. For example, a short nasal segment can be produced prior to a voiced stop in Spanish in order to facilitate the production of vocal fold vibration. More specifically, voicing can easily be extinguished in /b, d, g/ because the closure can cause the air pressure in the mouth to approach that in the lungs (below the vocal folds) which, for aerodynamic reasons, would cause the vocal folds to stop vibrating. The production of a preceding brief nasal before the voiced stop reduces the air pressure in the mouth (by channelling the air through the nose) so that the conditions for vocal fold vibration are once more met. Solé [138] shows how this type of implementational feature could have led to the historical development of prenasalised /^{m}b, ^{n}d, ^{ŋ}g/ from voiced stops /b, d, g/ in e.g. Austronesian, Papuan, and South American languages [137]. She also suggests that sound change may often involve just such implementational features, partly because their use varies so much both across and within speakers making it difficult for listeners to parse them with the source (the voiced stop).

iii. *Trading relationships*. In speech production, there are typically multiple acoustic cues for communicating a particular contrast. When two cues are in a trading relationship, then the strength of one cue can vary inversely with the other. For example, /p/ is distinguished from /b/ by a long voice-onset-time (VOT) and/or high-amplitude aspiration noise. In general, the cut-off at which listeners stop hearing /p/ and start to hear /b/ depends on a trading relationship between these cues: the longer the VOT, the lower the amplitude aspiration noise must be (and vice-versa) in order for the cut-off point to remain the same [126]. For Beddor [12], the development of a trading relationship is fundamental to explaining how certain types of coarticulation can evolve into sound change. To return to the VN sequence considered earlier, before the sound change takes hold, listeners factor out nasal coarticulation from the vowel, as described above. However, at some point from synchronic variation to sound change, nasalisation in the vowel and the following nasal consonant are presumed to enter into a trading relationship such that listeners no longer necessarily parse contextual vowel nasalisation with its source (a following N), but instead perceive nasalisation either from information in the vowel or from the following nasal consonant. It is this perceptual change from initially factoring out coarticulation to deploying it in a trading relationship which can begin to provide an explanation for why the source is often deleted as the sound change takes hold [56]. This is because, the more listeners depend on nasal coarticulation in the vowel (to identify nasalisation in a VN sequence), then the less they depend on acoustic information in the following nasal consonant (since the cues are in a trading relationship); thus the extreme case in which nasalisation is perceived entirely in the vowel gives rise increasingly to the perceptual extinction of the source, if the cues are in a perceptual trading relationship. There has been a long-standing puzzle in historical linguistics concerned with how a sound change such as the developing of vowel nasalisation (or indeed umlaut) could result from coarticulation if the source that gives rise to it is subsequently deleted—because if the source wanes then so too should the coarticulatory effect as a result of which the conditions for sound change to take place would no longer be met (see Janda [71] for a further discussion). The idea that the perceptual enhancement of coarticulation may be coupled with this weakening of the source if these cues are in a trading relationship may begin to provide an answer to this puzzle that is grounded in the mechanisms of perception (and how they are related to production).

iv. *Lexical frequency*. Statistical properties of the lexicon and in particular the frequency with which words occur in the language are for some [24, 28, 119, 120] central factors in explaining why sound change occurs. This issue touches upon a long-standing debate of whether, following the Neogrammarian principles [118], sound change spreads through all words of the lexicon at the same rate or whether instead it takes hold first in more frequent words [131, 154]. One of the main reasons why the association between word frequency and sound change is complex is because high frequency words show durational shortening and are often spatially reduced compared with low frequency words [38, 91, 161]. Thus it is not clear whether lexical frequency makes a contribution to sound change independently of a more casual speaking style which also tends to be temporally and spatially reduced. The recent study by Lin et al. [91] points, however, to a rather more direct role of lexical frequency in providing the conditions for sound change to occur: their ultrasound study shows that the tongue tip of /l/ is more likely to be lenited in more frequent words like *milk* than in less frequent words like *elk* but importantly without the /l/ being spatially more reduced due to changes in speaking style. Independently of this finding, an investigation by Zellou and Tamminga [164] showed more extensive nasal coarticulation in more frequent English words in the absence of any durational reduction. The importance of these studies lies in demonstrating therefore that lexical frequency is an independent factor that may predispose sound change to take place. Moreover, given the finding in Lin et al. [91], it seems plausible, as Zellou and Tamminga [164] suggest, that the increased nasal coarticulation in lexically more frequent words is also accompanied by a greater lenition of the nasal stop closure: this would imply that, in frequent words like *man*, there is extensive nasalisation in /a/ coupled with tongue tip lenition in the production of /n/. Thus lexically more frequent words may provide the conditions not just for increased coarticulation but also for greater lenition and the subsequent deletion of the source that gives rise to contextual nasalisation (leading to the type of sound change discussed above in which French /mɛ̃/ evolved from Latin *manus*).

v. *Hyper- and hypoarticulation*. According to Lindblom et al. [94], sound change derives from the adaptation of everyday speech communication to the needs of the listener. In this model, speech is produced with extensive spatial and duration reduction (hypoarticulated speech) based on the speaker's prediction that the listener can compensate for the resulting unclear speech by bringing to bear knowledge about the structure of the language. In hypoarticulated speech, the listener's attention is typically not focused on the signal (which often is lacking in clarity) but instead on the content (the semantics of the utterance). If exceptionally the listener processes the phonetic details of hypoarticulated speech, then new forms can be suggested that are added to the listener's lexicon: this is one of the main ways in which reductive sound change (such as 'chocolate' now produced with two syllables /tʃɒklət/) could be added to the lexicon. If coarticulation increases under hypoarticulated speech, then such a model would be able to explain many of the coarticulation-induced sound changes discussed earlier. However, it is not at all clear that coarticulation really does increase in a hypoarticulated speaking style (e.g. Matthies et al. [101], Bradlow [20]). The recent experiments in Harrington et al. [57, 59] and Siddins et al. [134] suggest that hypoarticulation does not magnify coarticulation in production, but it does degrade the listener's ability to factor out the influence of coarticulation: that is, under hypoarticulated speech, there is a more ambiguous relationship between coarticulation and the source that is responsible for coarticulation—just the condition according to which sound change should take place, according to Ohala's [111, 112] model.

jmh@phonetik.uni-muenchen.de

vi. *Mismatch between production and perception dynamics*. Harrington et al. [54] and Kleber et al. [82] investigated the relationship between the production and perception of coarticulation for the same speakers during a sound change in progress. The results in Kleber et al. [82] suggest that the association between the perception and production of coarticulation becomes unstable during an ongoing sound change, such that the two modalities are out of alignment (in those typically younger subjects who participate in the sound change) and in which changes to the coarticulatory relationships in perception lead those in production. These findings were predicted from Ohala's [112] model according to which sound change originates initially from the listener's failure to parse coarticulatory dynamics in accordance with the way in which they are produced.

Taken together, points i. and iv. can begin to provide a model for the conditions which predispose sound change to occur that is not functional or teleological (planned). Under a functional view, sound change occurs with the aim of maintaining or even enhancing meaning contrasts. Such a view is central to Martinet's [100] explanation of so-called vowel chain shifts in which the diachronic change in the position of a vowel can have a knock-on effect such that vowels push or pull each other around the vowel space. One of the most recent and striking examples of what seems to be a vowel chain shift is in New Zealand English in which in just the last 50–60 years the front vowels have rearranged themselves such that /a/ → /ɛ/ → /ɪ/ and in which /ɪ/ has centralised: for this reason, a New Zealand English production of *dad* sounds like *dead* and *desk* like *disk*, while *disk* has a vowel quality that is perceived to be not too dissimilar from /ə/ for English speakers of many other varieties [63, 98, 156]. A functional interpretation in Martinet's [100] terms would be that the raising of /a/ to /ɛ/ *causes* the /ɛ/ to raise to /ɪ/ *in order that* the contrasts can be maintained between the three vowels. A functional interpretation is implicit in the model of Lindblom et al. [94] in v. above because sound change arises out of a strategy *in order to* enhance or reduce contrasts based on a prediction of the listener's knowledge.

On the other hand, there may be enough in how the mechanics of speech production, speech perception and the structure of the lexicon are connected without the need to resort to functional explanations. For example, since variation is more likely in lexically frequent words (e.g. Aylett and Turk [3], Wright [161]), then the type of articulatory variation leading to a quantal acoustic change sketched in i. should also be more probable in lexically frequent words. A quantal acoustic change may be one of the factors that contributes to the greater tendency to unlink the coarticulatory variation from the source—to parse nasalisation with the vowel for example in VN sequences. Moreover, the source may be prone to disappear in lexically frequent words because, quite apart from the potential contribution of perceptual trading relationships to sound change (point iii.), the source is reduced (point iv. above) in frequent words and may therefore be less perceptible than in less frequent words. This would complete (or phonologise) the sound change by which in this example nasalisation is associated with the vowel together with deletion of the nasal consonant. Notice how all such principles draw upon naturally occurring phenomena in the production and perception of speech without the need to resort to a functional explanation.

It may also be possible to invoke the same machinery to explain why phonologisation can lead to enhancement [69]. Consider that the vowel /y/ which, in words like *Füße* ('feet'), developed from an /o/ or /u/ in /fotiz/ under the coarticulatory influence of the following /i/, has a much more front quality than would be predicted by synchronic vowel-to-vowel coarticulation. That is, Öhman's [110] study showed that the vowels influence each other but the tongue dorsum is not pushed as far forward as it is in modern German /y/ by the coarticulatory influence of $V_2 =$ /i/ on $V_1 =$ /u/ or /o/. Similarly, while there are phonetic reasons for the vowel to be shorter before a following voiceless than voiced consonant [27, 89], the magnitude of this shortening before voiceless (e.g. *bus*) compared with voiced (*buzz*) consonants in English is far greater than would be predicted by phonetic shortening alone (Ohala and Ohala [114]; see also Solé [136]). Kirby [78, 79] has recently developed a computational model in which phonologisation is an emergent consequence of a combination of precision loss and enhancement. More specifically, the model in Kirby [78] suggests that, as one cue for distinguishing between phonologically different categories wanes, another will be enhanced to ensure that the categories remain distinct. The cue that is enhanced is not necessarily the one that, as in Kingston and Diehl [77], combines with other cues in such a way to enhance an existing phonological contrast (e.g. lip-rounding with tongue backing in the case of /u/), but is instead whichever cue is likely to lead to the greatest probabilistic separation between phonological categories: in Kirby's computer simulation, the cue that comes to be enhanced depends on its acoustic effectiveness for separating between phonological categories, combined with the degree to which it is of use in distinguishing between items in the lexicon. The computer simulation is able to model speech data presented in Kirby [78, 79] showing how fundamental frequency has taken over from duration as the main cue in separating /CrV, CV, ChV/ in the Phnom Penh variety of Khmer.

It is tempting to conclude (as suggested by Kirby's model) that enhancement may be a consequence of phonologisation: speakers enhance the distinctions *because* they are functional i.e. provide an acoustic sharpening of the sounds that are involved in contrasting meaning. Alternatively, consider as discussed earlier that phonologisation often involves the attrition or deletion of the source (of e.g. the nasal consonant) that gives rise to coarticulation (e.g. the nasalisation in the preceding vowel). If as explained in iii. above, the source (nasal consonant) and coarticulatory effect (vowel nasalisation) are in a perceptual trading relationship, then source deletion implies an extreme or at least progressively increasing form of coarticulation, since by definition the two cues are inversely proportional (if they are in a trading relationship). Thus enhancement of coarticulation leading to phonologisation need not be functional, but may be the outcome of what happens when the coarticulatory effect and source enter into a (progressively one-sided) perceptual trading relationship, leading to enhanced coarticulation combined with source attrition.

6.2 Sound Change and Experience

Developments in the last 10–15 years in the episodic or exemplar model of speech perception and production [44, 73, 109, 120] give much greater emphasis to the way in which differences between speakers in their learned phonetic knowledge can contribute to the conditions for sound change to occur. Models of sound change—and indeed more generally speech communication—within an exemplar paradigm are necessarily social: from this point of view, it makes little sense to develop models of sound change that are formulated independently of speaker or listener variation. An exemplar model applied to sound change goes some way towards providing a common framework for models such as Ohala's [111, 112] concerned with the conditions that give rise to sound change and those within the sociolinguistic tradition [85] whose central focus is more on how sound change spreads within a community.

The central idea in an exemplar model of speech is that listeners store in memory episodes i.e. unsegmented auditory gestalts of the words that they hear in a high dimensional perceptual space. Phonological units—that is the abstract units whose permutations function to distinguish between words—emerge from regions of high density in this multidimensional space [121, 122]. Since, for example, all words beginning with /t/ share some basic acoustic properties (of e.g. a rising spectrum in the burst, a second formant frequency that converges around a locus close to 1800 Hz etc.), then the unsegmented auditory traces of such words converge around the (auditory transformation) of these acoustic characteristics: it is such dense regions of intersection that provide the conditions for abstraction i.e. for learning that there is an abstract category such as /t/. Moreover, it is not just phonological information that emerges from stored episodes but also dialect and sociolinguistic information, as well as other speaker characteristics such as gender and age (see Docherty and Foulkes [31] for a further discussion).

The following are some of the most important ways in which exemplar theory has contributed to, or could provide an explanation for, how synchronic variation and sound change are connected.

i. *Coarticulation*. As discussed earlier, Ohala's model suggests that the conditions for the occurrence of sound change are met if listeners misperceive coarticulatory relationships. One of the reasons why they might do so is because the speech production-perception link is non-linear: that is, there are certain kinds of coarticulatory overlap in speech production that are poorly or ambiguously transmitted in speech perception. However, exemplar theory provides at least another reason: if the association between speech and phonological units is idiosyncratic at the level of the individual, then listeners even of the same speaking community may not agree on how to parse coarticulation. For this reason, and as Baker et al. [7] suggest, variation between speakers may make normalising for the effects of coarticulation more difficult. There is, moreover, evidence that listeners parse coarticulation in the same signal differently. For example, Fowler and Brown [37] and Beddor [12, 13] have shown how the extent to which listeners parse nasal coarticulation in the vowel with the following nasal in VN sequences is listener specific. Stevens and Reubold [143] demonstrate the listener-specific way in which pre-aspiration is parsed either with the preceding vowel or with the following geminate in Italian. Yu [162] has extended such findings to show that parsing coarticulation with the source that gives rise to it is affected by the listener's personality and social profile including the extent of autistic-like traits. Harrington et al. [54] and Kleber et al. [82] have both demonstrated perceived coarticulation differences at the group level (see also Kataoka [74]): for the sound change in progress by which /u/ has become fronted under the coarticulatory influence of preceding fronting consonants, older and younger listeners were shown to differ in the extent to which they normalise for coarticulation (Fig. 6.2). Baker et al. [7] have proposed that the conditions for sound change to occur depend on some speakers who produce coarticulation in a particularly exaggerated way—it is these speakers that may be more likely to be imitated (see iv. below) leading to the sound change to be propagated through the community. Baker et al. [7] reason that sound change may be rare (in relation to the ubiquity of synchronic variation) not just because listeners compensate in perception so effectively for context in the way suggested by Ohala, but also because of the scarcity of individuals who exaggerate coarticulation to such an extreme degree. Another reason why sound change may be rare is because only a small number of listeners may respond to and imitate the novel variants that occur in producing exaggerated coarticulation [47].

ii. *Lexical frequency*. More frequent words necessarily give rise to more episodes because by definition we hear them more often. For this reason, the association between phonological categories and speech communication is skewed by statistical properties of the lexicon in an exemplar model. An analysis by Hay and Foulkes [61] of archival recordings of New Zealand English has shown that the progression of the reductive sound change of domain-final /t/-deletion is faster in lexically frequent words. Reubold and Harrington [127] provide some preliminary evidence of an association between sound change and lexical frequency in the same individual. Part of their study was concerned with a longitudinal analysis of the broadcaster Alistair Cooke over a 70-year period. They showed that, having acquired aspects of General American after emigrating to the United States in the 1930s, his accent in later life was becoming again closer to the Received Pronunciation that he produced prior to emigrating. They also provided evidence that this reversion to Received Pronunciation (involving a backing of the vowel in the lexical set BATH) had taken place at a faster rate in lexically frequent than infrequent words. On the other hand, while Labov [88] finds no effect of lexical frequency in vowel chain shifts, Hay et al. [63] show that vowel chain shifts in New Zealand English are led by lexically *infrequent* words because, according to their theory, infrequent words are less likely to have an impact on long-term memory when they occur in regions of vowel overlap.

iii. *Incrementation*. For the Neogrammarians, regular sound change was incremental and not perceptible. In more recent times, the issue of whether regular sound change is incremental or not is more controversial. Thus for some (e.g. Ohala [112], Baker et al. [7]) regular sound change involves an abrupt change between different phonetic variants while for others [15, 56, 83, 102], aspects of regular sound change can be incremental. Irrespective of the controversy, incrementation is predicted by exemplar theory to the extent that sound change comes about through updating a cloud of remembered episodes. It is because the cloud itself is a statistical generalisation over a very large number of episodes that a sound change will initially have only a very small impact, in much the same way that outliers have a small influence on statistical parameters like the mean and variance of a tightly fitting Gaussian probability distribution. But as more and more such outliers accumulate, then the distribution will shift (but incrementally).

iv. The puzzle of the *actuation of sound change* (Weinreich et al. [158]; see also Stevens and Harrington [142] for a recent review) which is concerned with why sound change should take hold at a particular time in one dialect or language but not in another has an explanation in terms of exemplar theory in which the association between episodes and categories, being based on usage and experience, is probabilistic. Independently of sound change, exemplar theory predicts that languages or dialects are likely to differ for similar sets of phonological contrasts: for example, although very many languages contrast voiced /b, d, g/ and voiceless /p, t, k/ stops, there are no known two languages that do so in quite the same way [123]. Just this is to be expected if phonological categories emerge stochastically from speech signals transmitted between speakers and hearers. Analogously, the unstable conditions of the kind reviewed in Sect. 6.1. that could lead to sound change may obtain in one dialect or language but not another, given the presumed probabilistic association between speech signals and phonological categories.

v. *Changes over the lifespan*. This issue (which is discussed in more detail under Sect. 6.4) is concerned with the way in which adults' pronunciation has been shown to change incrementally often over several decades (e.g. Harrington [50], Quené [125], Sankoff and Blondeau [130]) in the direction of changes that have been taking place in the community [53]. The prediction from exemplar theory is that the shift should be dependent on experience. For example, assuming additionally a model in which first and second language acquisition share the same phonetic space, an exemplar model predicts findings such as those in e.g. Sancier and Fowler [129] of a change in a bilingual Portuguese-English adult's productions of American English plosives in the direction of the (shorter) voice onset time of Portuguese, after spending several months in a Portuguese speaking environment in Brazil.

vi. *Category broadening and narrowing*. If, as exemplar theory suggests, categories are derived from absorbing episodes of speech signals that a listener encounters, then categories will evolve into ones that are infinitely broad (due to variation) as a result of which there would be a complete collapse of contrast. The mechanism by which category distinction and stability are maintained could be some form of averaging of episodes in speech production [120]: that is, all episodes are absorbed in perception, but production is based on an aggregate of episodes. This aggregate would create its own episode that, for reasons of statistical sampling, would typically be close to the centre of the distribution. In this way, production, if based on an aggregate, would strengthen the mean of the category; this would, in turn, counteract a category's unlimited broadening. Thus, any phonological category is acted upon by two forces: one that creates a broadening of the category by the uptake of new episodes that are at the probabilistic edges of the distribution; and one that creates a contraction towards the centre through averaging episodes in speech production. It is when one of the edges comes into contact with the distribution from another category that the potential for sound change exists. This is because the edge tokens (for example /e/ vowels with a particularly high F1) will be absorbed into the other category (e.g. /a/). Since tokens are lost at one of the edges, then the force pulling the category in that direction will be weakened, and as a consequence the category will shift slightly in the *opposite* direction (/e/ would shift towards lower F1 values in this example). This change in the balance between opposing forces is simulated in a computational model by Blevins and Wedel [17]. They use the model to explain why the sounds of words whose meaning is not otherwise resolved by context tend not to merge diachronically in just the conditions that would give rise to a loss of contrast i.e. homophony (see their paper for numerous examples of anti-homophony from various languages). Such a model in which a shift in the balance between the strengthening force at the centre and the broadening force at the edge causing categories to repel each other when they are in close proximity could form the basis of explaining vowel push-chains of the kind that were detailed earlier for New Zealand English.

vii. *Sound change through not updating categories*. Exemplar theory suggests that sound change can come about if phonological categories are either not, or only selectively, updated by episodes. In Silverman's [135] analysis of the Mexican language Trique, a sound change has developed by which a /ug, ud/ contrast has evolved into /ug^{w}, ud/ i.e. with lip-rounding on /g/ but not on /d/. The lip-rounding evidently originates from the lip-rounded vowel /u/, but the issue is why /g/ but not /d/ should become rounded. The crucial insight here is that lip-rounding and /g/ are in a sense acoustically additive so that a lip-rounded /g^{w}/ is *further away* acoustically and perceptually from /d/ (e.g. Halle et al. [49]), whereas a lip-rounded /d^{w}/ would cause a lowering of its burst's spectral centre of gravity resulting in an acoustic shift *towards* /g/ (given that /d/ is distinguished from /g/ by high frequency energy in the spectrum; see Harrington [51] for further details). The evolution of this sound change is modelled in terms of rejecting exemplars in regions of ambiguity between categories. Thus, according to this model, a listener is more likely to reject unrounded /ug/ tokens (because these are closer to /ud/ than is /g^{w}/) and is similarly more likely to reject /ud^{w}/ tokens (because these are closer to the /g/ distribution than to /d/). In this way, a gradual bias is introduced over successive generations of speakers by which the progressive rejection of /ug/ and /ud^{w}/ results in the emergence of an acoustically enhanced contrast /g^{w}, d/. Notice how this is enhancement without teleology: that is, the progressive emergence of the more distinct /g^{w}, d/ is a natural consequence of rejecting tokens i.e. not updating phonological categories in regions of ambiguity. Similarly, Hay et al. [63] show how a vowel push-chain in New Zealand English (of the kind described in i.) can emerge through not updating categories in regions of ambiguity i.e. ones in which the meaning is compromised. A computer model based on a similar principle is discussed in Garrett and Johnson [42] who also note that not updating categories with ambiguous tokens is consistent with Labov's [85] idea that misunderstood tokens are not used in speech production.

Boersma and Hamann [19] are also concerned with developing a model of sound change in which phonological distinctiveness emerges iteratively through non-purposeful interactions between teachers and learners. They are critical of the mechanisms in (vi., vii.) to maintain stability by having to resort to concepts such as an aggregate across a large number of exemplars in production [120]. Their model adopts by contrast from optimality theory the idea that phonological categories are associated with the speech signal by means of (some 322) constraint-ranked auditory filter bands. The interesting aspect of this model is that stability (i.e. no change) is a compromise between a minimisation of articulatory effort and minimisation of perceptual confusion (which is reminiscent of Lindblom et al. [92] model discussed earlier): sound change is likely to come about when these articulatory and auditory minimisations are not balanced. However, this model has at least as many and perhaps more artificial mechanisms than those proposed within the exemplar paradigm. These include the (so far undemonstrated) idea that children learn an OT-like constraint ranking by bringing into contact their correct knowledge of the underlying phonological structure of words with adult-like acoustic distributions of the phonological categories (a position which Boersma and Hamann admit is unrealistic); and that categories that are auditorily more peripheral are harder to produce (see also Kirby [78] for a similar criticism).

6.3 Sound Change and First Language Acquisition

The idea that there might be direct parallels between language acquisition and sound change has a long history [45, 70, 118, 146] and was taken up in many of the generative and natural [145] phonological frameworks of more recent times (e.g. Lightfoot [90], Kiparsky [80]—see Foulkes and Vihman [34], Beckman [11]; and Diessel [30] for a review). But there are, as shown in Vihman [153] and more recently Foulkes and Vihman [34], serious difficulties with any assumption of such a direct link between the errors produced by children and the forces that give rise to sound change. Firstly, the earlier assumptions about the relationship between acquisition and change are based on treating children as a homogeneous entity when in fact there are differences in the rate of acquisition that can be linked to the development of the lexicon (e.g. Beckman et al. [10], Munson et al. [103]). Secondly, and based on auditory analyses of children's misarticulations in five different languages, Vihman's [153] study shows that many of the typical misarticulations produced by children such as consonant harmony, the simplification of consonant clusters to single consonants, and the shortening of long words are quite uncommon in sound change. Thirdly (and most importantly) it is not enough, as Foulkes and Vihman [34] argue, just to list that there may be parallels between sound change and misarticulation by children (which in any case has not been demonstrated): it is instead necessary to understand the cognitive and social mechanisms by which children's misarticulations or misinterpretations of the speech signal may evolve into diachronic change.

A more fruitful approach towards understanding the connections between child language acquisition and sound change has recently been presented in Beckman et al. [11] who seek to demonstrate that the two types of sound change that are commonly referred to in the sociolinguistics literature as 'from below' and 'from above' the level of awareness interact differently with patterns of speech acquisition in the Seoul variety of Korean and in two varieties of Mandarin. The distinction between sound change 'from above' and 'from below' [85, 87] is not especially well defined, partly because what is above or below awareness or consciousness is not easily integrated with cognitive models of speech processing, and also because the distinction is confounded with social class (changes from a lower-middle class usually being from below; changes from a prestigious class, from above—see Wardhaugh and Fuller [155] for a further discussion). Nevertheless, there is a rough correspondence between sound change from below and the regular (as opposed to analogic) changes that have their origins in the processes of coarticulation and lenition, although sound change from below typically also encompasses a sociolinguistic dimension of less prestigious, more working-class speech that is often found in a local vernacular. Sound changes from below like those due to coarticulation are incremental in that they progress gradually across generations [87]. By contrast sound change from above—which is often brought about by contact between speakers of different dialects or languages—more typically involves an abrupt change from one phoneme or category to another and may be driven by a more prestigious or powerful group of individuals. A possible example of a sound change from above might be the increasing use in Standard Southern British English of high-rising-terminals in which the intonation rises in declarative sentences: there is no sense in which this change is incremental (it was not the case that falling intonation gradually evolved via mid-level intonation into rising intonation over a number of years), it is arguably brought about through dialect contact (with North American and/or Australian varieties) and it is a change that listeners have commented upon (and so cannot be, in Labov's terminology, below the level of consciousness). Notice as well that such a sound change is—like many sound changes from above—idiosyncratic in that, unlike the coarticulation-based sound changes discussed earlier, it is not usually found across multiple languages.

Labov [84, 87] has also argued that sound change from below is often led by women: the association with language acquisition is their typically greater involvement than men as caregivers. For example, Labov [84] proposes that many sound changes from below originate in the lower-middle classes; that for these classes there is the greatest differentiation between male and female speakers (especially since female speakers are most likely to style shift and copy pronunciations of the upper classes in a more formal speaking style); and that consequently, since women are the primary caregivers, infants' speech will be influenced more by women than by men (as a result of which sound change from below is often led by women). This view about which gender leads sound change is, however, complicated by more recent findings showing that speaking style in child-directed speech is differently affected by whether caregivers are talking to infant boys or girls (e.g. Foulkes et al. [35]).

Beckman et al.'s [11] study has made advances to our understanding of how the progress of sound change and language acquisition might be connected. They do so through analyses of adult data over a 60 year period combined with perceptual studies of a sound change by which in Seoul Korean the distinction between lax versus aspirated stops is increasingly cued not as in earlier times by voice onset time, but by fundamental frequency differences. They argue that the change is 'from below': this is because a similar type of sound change is found in many other languages; and, as is typical for a sound change from below, it has progressed at a faster rate in women. They then show that infants are not as advanced in this sound change: that is, infants make more use of VOT than younger women. This comes about, Beckman et al. [11] reason, because their caregivers are from an older generation. Following Hockett [66], it is only when children at a slightly older age mix with peers and other social groups that they will not just imitate but incrementally advance the change further still (see also Kerswill [76]). The other finding in Beckman et al. [11] is that the association with language acquisition is very different for the sound change from above by which a three-way fricative contrast has developed in the Sōngyuán variety under the influence of standard Mandarin. Importantly, children of the Sōngyuán variety do not show any incremental progression but instead copy categorically the three-way fricative contrast that does not exist for adult speakers of Sōngyuán.

A different approach to the relationship between sound change and acquisition that relates rather more directly to Ohala's model (see also Greenlee and Ohala [46]) is explored in Kleber and Peters [81]: their concern is to test whether as less experienced users of the language, children are more likely to have difficulty normalising in perception for context effects such as coarticulation. If this is so, then there may be a greater potential for children to fail to attribute a coarticulatory effect to the source that gives rise to it, thus providing the conditions for sound change to occur. A study by Nittrouer and Studdert-Kennedy [108] provided some evidence that adults normalise to a greater extent for the coarticulatory effects of vowel context on preceding fricatives than children; and that there was also a greater degree of adult-like normalisation in 7-year old compared with 3-year old children. Consistently with this study, Kleber and Peters [81] have found that children are much more variable in perception than adults in normalising for the effects of context. A subsequent perception study by Harrington et al. [58] compared young first language German children and adults on the extent to which they normalised for the coarticulatory influences of /p_p/ and /t_t/ contexts on German high front /ʊ/ (back) and /ʏ/ (front) vowels. Their results showed that the distance between the decision boundaries on an /ʊ-ʏ/ continuum were closer together for the children than for adults which suggests that children might have been less sensitive to the perceptual influence of the consonantal context on the vowel. However, the interpretation of adult-child differences in Harrington et al. [58] is not that children normalise less for coarticulation, but instead that they are less certain than adults about phonological categorisation, in a way that is analogous to the greater uncertainty when adult listeners categorise hypoarticulated speech signals [57, 59].

6.4 Sound Change and Imitation

The last few years have seen a focus on spontaneous imitation in speech (Delvaux and Soquet [29], Nielsen [106], Pardo et al. [116, 117], Yu et al. [163] to mention but a few) and some attempts to consider the implications of these findings for models of sound change [6, 42, 52]. The methodology for demonstrating imitation typically involves comparing subjects' speech production before and after they have performed a task such as listening to or shadowing another speaker; imitation has also been measured and demonstrated in terms of the degree of convergence between speakers in a conversation [116]. In general, two main findings have emerged from this research. The first is that imitation can take place without any social motivation to do so. This is so, for example, in Nielsen's [106] study in which imitation is measured from the production of isolated words. This type of spontaneous imitation may be part of a more general tendency for individuals to coordinate their actions in space and time [40, 132] that can give rise to alignment of many different kinds—for example of body movements [133] and of syntactic structures [43]. Secondly, imitation is by no means automatic and inevitable for all speakers [163] and can vary depending on the speaking situations [116]. In addition, imitation can be constrained by the social context including how unusual [6] or attractive [5] the interlocutor's voice is perceived to be, as well as the attitudes of the speaker towards the interlocutor's national identity [4].

In the analyses of the Christmas broadcasts produced annually by Queen Elizabeth II, Harrington et al. [53] showed that the Queen's 1950s vowels had shifted over 30 years towards those of a mainstream Standard Southern British (SSB) more typically produced by the middle classes. The progression which was found for several vowels, was gradual and in the direction of sound change that had been taking place to the standard accent of England over a 50 year period; it was also away from an aristocratic form of the Queen's 1950s variety (referred to by Wells [159] as U- or upper-crust RP) towards a more middle-class variety of SSB. Moreover, the shift was partial such that the Queen's 1980s vowels were generally at intermediary locations between those from the 1950s and those that were more typical of the five analysed mainstream SSB speakers in Harrington et al. [53] who had been recorded in the 1980s.

It seems quite probable that these shifts to the Queen's vowels have been brought about because of some form of imitation or convergence in dialogue. As discussed in Harrington [56], the 1960s and 1970s saw a rise of the lower-middle and middle classes into positions of power, with the result that the Queen is likely to have come into increasing contact in the intervening decades with speakers of a more middle class variety. Indeed, the prime ministers to which the Queen gives a weekly audience were in the 1960s and 1970s (James Callaghan, Edward Heath, Margaret Thatcher, Harold Wilson) generally of more humble origins than their predecessors in the

1950s and (with the possible exception of Heath) with accents that were removed from U-RP of the 1950s.

The question that must be considered is why the Queen should shift her accent towards those speakers with a less aristocratic, more mainstream form of Standard Southern British which would include not just her prime ministers but quite possibly also staff who arguably might have been increasingly likely to produce a less aristocratic form of RP after the 1950s. Perhaps it was because the Queen engaged in style-shifting [62] towards an accent that typified a more egalitarian society that was shaping England in the late 1960s and 1970s. But the reason why this interpretation is unlikely is because the sound change in these broadcasts is gradual. Moreover, the changes to the Queen's vowels are slow (some 10 Hz per annum in the fronting of /u/) which would have repeatedly required quite accurate style-shifting of a few Hz per annum over the thirty year period [52]. A more plausible explanation is that personal accent is incrementally influenced over a long period of time through daily or regular spoken interaction with speakers from another dialect group or indeed from another language [129]. Incrementation is, as explained earlier, also predicted by exemplar theory: the accumulation over time of episodes of speech signals from speakers of other dialect groups should shift incrementally the statistical distributions of a person's phonological categories: that is, interacting with and listening to speakers over 2–3 decades who have a phonetically more fronted /u/ has fronted incrementally the Queen's 1950s /u/ towards a form intermediate between U-RP of the 1950s and SSB of the 1980s.

The positions of the Queen's 1980s vowels between those of her 1950s vowels and those of the five mainstream SSB speakers recorded in the 1980s is predicted by a dialect mixture model in which there has been a blending or averaging of the aristocratic, conservative accent of the 1950s with the more middle-class, mainstream SSB accent of the 1980s.

The above interpretation of incrementation through interaction is also entirely consistent with the model of dialect mixture that is developed by Trudgill [151] in his analyses of New Zealand English. Trudgill argues that the creation of this accent has been a deterministic function of a levelling of the different dialects of the speakers who first came to New Zealand in the 19th century: that is, New Zealand English is not the result of the need for New Zealanders to establish group membership and identities that are unique to their country. For Trudgill [151], the evolution of New Zealand English as a consequence of dialect mixture is due to an 'innate tendency to behavioral coordination' (p. 252) among the interlocutors, a process that is the consequence of interaction. Building on an earlier idea from Bloomfield [18], he notes that this model of dialect mixture is in agreement with Labov's [86] view that in many cases the diffusion of linguistic change can be explained through communication density i.e. as a consequence of the speakers interacting with each other. Importantly, Labov [86] notes that there need not be any motivating social force behind such changes which can be mechanical and inevitable and in which social evaluation plays only a very minor role (Labov [86], p. 19–20). Trudgill's [151] idea of new dialect formation via non-social determinism is a further development of his earlier gravity model in which the spread of a linguistic innovation depends on a combination of geographical distance and population density. Trudgill's [148] gravity model was used to explain how an innovation such as /h/-dropping spreads in England from city to city while skipping over less populated rural areas. As described in further detail by Wolfram and Schilling-Estes [160], the gravity model was subsequently modified to take account of other findings [149] suggesting the likelihood of change is greater when the two dialects are phonetically similar to each other.

This type of dialect averaging has recently been demonstrated for a completely different set of data by Bukmaier et al. [22] in a study concerned with dialect levelling due to contact with a standard variety. The focus of their study is the Augsburg variety of German which, in contrast to standard German, neutralises the /s-ʃ/ contrast in certain pre-consonantal contexts: thus standard German /mɪst, mɪʃt / (*misst*/*mischt*; 'measures'/'mixes') are both /mɪʃt / in Augsburg German. Now it might be supposed that any change in the direction of the standard is categorical by which younger speakers typically produce more /s/ in words like *misst* than their older counterparts. But this is not what the instrumental analysis showed: instead Bukmaier et al. [22] found that younger Augsburg subjects were intermediate between older Augsburg and standard German subjects in the fricative of words like *misst*. This was so in both production and perception. Thus their produced forms had all the characteristics of a sibilant fricative that was intermediate between /s/ and /ʃ/; and in perception, they were also intermediate between the poor ability of older Augsburg and sharp discrimination of standard listeners in distinguishing between word pairs like /fɛɐ̯ˈmɪstə, fɛɐ̯ˈmɪʃtə/ (*vermisste*/*vermischte*; 'missed'/'mingled'). There was a similar finding in an apparent time study comparing speakers of East Franconian German with standard German speakers [56, 105]. In East Franconian, as in many other varieties of German, the post-vocalic voicing contrast is neutralised towards the voiced variety: thus *leiten*/*leiden*, /laɪtn, laɪdn/ ('to lead'/ 'to suffer') are both /laɪdn/ in the East Franconian variety. However, younger East Franconians were found to produce and perceive a contrast that was intermediate between those of older East Franconians and standard German subjects. Such intermediate positions for both Augsburg and East Franconian German during a sound change in progress due to dialect contact are exactly what is predicted by Trudgill's [151] dialect mixture model.

Trudgill [150] also shows that the first two generations of young children were likely to have been the prime instigators of the dialect mixture leading to the development of New Zealand English; he also suggests [152] that children may be especially prone to the type of interactional alignment described above. Children's accents are of course very malleable: the accent of a child rapidly shifts towards that of its peers and away from that of its parents, which is especially noticeable if they and their families are recent migrants to a different dialect region, as Chambers [26] and others have shown. Recently, Nielsen [107] showed that there was more imitation of (a lengthened) voice onset time in stops by children—both pre-schoolers and those aged 8–9 years than by adults. Consider in addition Babel's [6] recent finding of a predisposition to imitate novel voices. Perhaps then one of the reasons why children may be predisposed to imitate is because of the large number of novel voices that they encounter as the child's social network typically expands beyond that of the family and caretakers. The imitation may be especially likely because, having previously encountered only a few voices and having relatively limited speaking experience, the statistical distributions of their emerging phonological categories would still be quite unstable and therefore strongly affected by novel voices and pronunciations (see Stanford and Kenny [139] for similar views).

On the other hand, computational simulations using agent-based modelling suggest that imitation and non-social determinism—which Trudgill [151] argues are central factors in new dialect formation—may not be sufficient to explain how sound [8] or language [124] change propagate through the community. Agent-based models apply principles from statistical physics to social dynamics in order to understand how local interaction between individuals—represented by agents who are inter-connected in a network—can bring about global (community) changes [25]. The agent-based social network model in Fagyal et al. [33] was designed to test various theories about how the inter-connectedness of individuals is related to linguistic change. Their general conclusion is that language change is propagated principally by so-called leaders that is those with many connections to other individuals and that linguistic change will only spread through the community if there are both leaders and 'loners' i.e. those with far fewer social connections. Their model also provides some support for the idea that prestige drives language change. In Pierrehumbert et al.'s [124] agent-based computational model, an individual's choice between alternative linguistic categories is based on three main factors: the community norm modelled as the aggregated input from the other speakers with which the individual is linked in the network; stored knowledge consisting of a bias factor that governs whether a given individual prefers conservative or novel forms; and finally a variable (designated as 'temperature') that controls the degree to which individuals randomly or categorically choose between alternatives (i.e. their model incorporates the idea that individuals select between alternatives with varying degrees of probability). Compatibly with Fagyal et al. [33], their simulations show that a mixture of individuals who are innovators (prone to adopt a form that is not consistent with community norms) and who resist change are prerequisites for language change. But in contrast to Fagyal et al. [33], highly connected individuals tend not to be innovative in Pierrehumbert et al.'s [124] simulations because their output is so strongly influenced by the large input that they receive from other more conservative individuals with which they are connected. Change is instead more likely to radiate from individuals in their model who are innovative, connected to other individuals who are similarly biased towards innovation, but who are only average or below average as far as the total number of connections to other individuals is concerned. Pierrehumbert et al. [124] use this model to suggest that sound change originates from closely-knit communities whose speakers share innovations and also to argue against the idea that prestige is a driving force for sound change. One of the major innovations in this model is that a probabilistic (rather than a binary) choice between categories or variants is one of the components in their model for language change to spread through the community.

6.5 Concluding Comments

Sound change is fascinating and scientifically tractable. There are speech signals on the one hand and there are categorical changes (*bird* < *bridde*) on the other and there is some form of cognitive, physiological, and perceptual machinery linking the two. The task is to assemble many different forms of empirical evidence—ranging across etymological reconstructions of changing sound patterns between ancient and modern languages, the social aspects of speech, physiological analyses of movement and their perceptual consequences—to find out what form the machinery takes in order to explain how signals and sound change are connected. This task is in turn illuminating for the most challenging task in the speech sciences: how the categories of speech—the consonants and vowels out of which words are constructed—and signals in human speech communication and the machine derivative thereof are interrelated. This is fundamental knowledge for both human and machine processing.

The review of the current state of the literature has shown that some themes for explaining the operation of sound change recur in several studies often across different disciplines. Some of the most important of these are as follows.

i. *Coarticulation* or more generally the dichotomy between the serial order of phonological categories and the interleaved, shingled movements in the speech signal remains fundamental to many different types of sound change (assimilation, dissimilation, metathesis, phonologisation) that recur through different languages.
ii. *Perception and its relationship to speech production*. Many different studies have shown how non-linearity between the two modalities can provide the conditions for sound change to take place: this is so when coarticulatory dynamics are misperceived which may also provide the basis for explaining the often asymmetric direction of sound change. In addition, coarticulation leading to sound change may become perceptually salient when articulatory incrementation pushes the acoustic signal across a quantal boundary. Perceptual trading relationships could be important in understanding phonologisation and also why phonologisation seems to lead to enhancement.
iii. *Hypoarticulation* which is the condition under which the speech signal is degraded often in relation to its predictability from context is important for understanding sound change, not just from the perspective that many types of sound change are reductive (involving e.g. consonant lenitions—e.g. Bybee [23]) but also because hypoarticulation combined with a possible degradation of the coarticulatory source may obscure coarticulatory parsing in perception [58, 59]. Hypoarticulation accounts also emphasise how stored (top-down) knowledge—such as lexical frequency may be implicated in sound change.
iv. *Experience* (exemplar) based models of speech have provided numerous additional ways and also an appropriate metalanguage for understanding the operation of sound change. Experience-based models can explain how speaker variation—both at the individual and at the group level—can create the conditions for sound change to occur. The perception-production feedback loop [157]

inherent in exemplar theory coupled with findings from imitation can explain the occurrence of sound change over the lifespan and the emergence across generations of a mixture when two dialects come into contact. This perception-production feedback loop explains the effects of lexical frequency; it also provides the mechanism for (and predicts) that regular sound change is incremental.

Finally, and as discussed in Blevins and Wedel [17], in much the same way that Darwin's theory of natural selection has led to an understanding of how purposeless interactions at the level of the individual can give rise to what appears to be a purposeful population change, so too can cumulative variation at a number of nested and interlocking levels between individual speakers and a community, and between speech signals and the lexicon, push the sounds of the language between stable and unstable states. This metaphor also provides a way of understanding sound change without having to invoke the untestable (or at least so far undemonstrated) idea that sound change is started by any individual (or pair of individuals as in Ohala's [112] model) and then spreads through the community. It also implies that the sharp division that is made by many between the so-called origin of sound change and its spread (e.g. Ohala [112], Baker et al. [7], Janda and Joseph [72]) is a fallacy. The sounds of language are in a constant state of flux i.e. there is no point at which sound change is not taking place (and so no fixed point at which sound change starts). There are instead multiple conditions that can create instabilities that lead to categorical change. Understanding these through the wide range of theories and empirical techniques that have been reviewed in this paper remains an exciting challenge in the future.

Acknowledgments This research was supported by European Research Council grant number 295573 'Sound change and the acquisition of speech' (2012–2017). We are very grateful to Daniela Müller, two anonymous reviewers, and the editors for very helpful comments.

References

1. Abrego-Collier C (2013) Liquid dissimilation as listener hypocorrection. In: Proceedings of the 37th annual meeting of the Berkeley linguistics society, 3–17
2. Alderete J, Frisch S (2006) Dissimilation in grammar and the lexicon. In: de Lacy P (ed) The cambridge handbook of phonology. Cambridge University Press, Cambridge
3. Aylett M, Turk A (2004) The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Lang Speech 47:31–56
4. Babel M (2010) Dialect divergence and convergence in New Zealand English. Lang Soc 39:437–456
5. Babel M (2012) Evidence for phonetic and social selectivity in spontaneous phonetic imitation. J Phon 40:177–189
6. Babel M, McGuire G, Walters S, Nicholls A (2014) Novelty and social preference in phonetic accommodation. Lab Phonol 5:123–150
7. Baker A, Archangeli D, Mielke J (2011) Variability in American English s-retraction suggests a solution to the actuation problem. Lang Var Chang 23:347–374

8. Baxter G, Blythe R, Croft W, McKane A (2009) Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English. Lang Var Chang 21:257–293
9. Beckman M, Edwards J (1994) Articulatory evidence for differentiating stress categories. In: Keating P (ed) Papers in laboratory phonology III. Cambridge University Press, Cambridge, pp 7–33
10. Beckman M, Munson B, Edwards JR (2007) Vocabulary growth and the developmental expansion of types of phonological knowledge. In: Hualde J, Cole J (eds) Papers in laboratory phonology IX. Cambridge University Press, Cambridge, pp 241–264
11. Beckman M, Li F, Kong E, Edwards J (2014) Aligning the timelines of phonological acquisition and change. Lab Phonol 5:151–194
12. Beddor P (2009) A coarticulatory path to sound change. Language 85:785–821
13. Beddor P (2012) Perception grammars and sound change. In: Solé M-J, Recasens D (eds) The initiation of sound change. Perception, production, and social factors. John Benjamin, Amsterdam
14. Beddor PS, Krakow RA (1999) Perception of coarticulatory nasalization by speakers of English and Thai: evidence for partial compensation. J Acoust Soc Am 106:2868–2887
15. Blevins J (2004) Evolutionary phonology: the emergence of sound patterns. Cambridge University Press, Cambridge
16. Blevins J, Garrett A (2004) The evolution of metathesis. In: Hayes B, Kirchner R, Steriade D (eds) Phonetically-based phonology. Cambridge University Press, Cambridge, pp 117–156
17. Blevins J, Wedel A (2009) Inhibited sound change: an evolutionary approach to lexical competition. Diachronica 26:143–183
18. Bloomfield L (1933) Language. Holt, New York
19. Boersma P, Hamann S (2008) The evolution of auditory dispersion in bidirectional constraint grammars. Phonology 25:217–270
20. Bradlow AR (2002) Confluent talker-and listener-oriented forces in clear speech production. In: Gussenhoven C, Warner N (eds) Papers in laboratory phonology VII. Mouton de Gruyter, New York, pp 241–273
21. Browman C, Goldstein L (1992) Articulatory phonology: an overview. Phonetica 49:155–180
22. Bukmaier V, Harrington J, Kleber F (2014) An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: evidence for a gradient sound change. Front. Psychol 5:1–12
23. Bybee J (2001) Phonology and language use. Cambridge University Press, Cambridge
24. Bybee J (2002) Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. Lang Va Chang 14:261–290
25. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. Rev Mod Phys 81:591–646
26. Chambers J (1992) Dialect acquisition. Language 68:673–705
27. Chen M (1970) Vowel length variation as a function of the voicing of the consonant environment. Phonetica 22:129–159
28. Chen M, Wang W (1975) Sound change: actuation and implementation. Language 51:255–281
29. Delvaux V, Soquet A (2007) The influence of ambient speech on adult speech productions through unintentional imitation. Phonetica 64:145–173
30. Diessel H (2012) Language change and language acquisition. In: Bergs A, Brinton L (eds) Historical linguistics of English: an international handbook, vol 2. Mouton de Gruyter, Berlin, pp 1599–1613
31. Docherty G, Foulkes P (2014) An evaluation of usage-based approaches to the modelling of sociophonetic variability. Lingua 142:42–56
32. Egurtzegi A (2014) Towards a phonetically grounded diachronic phonology of Basque. PhD dissertation, University of the Basque Country
33. Fagyal Z, Escobar A, Swarup S, Gasser L, Lakkaraju K (2010) Centers and periph- eries: network roles in language change. Lingua 120:2061–2079
34. Foulkes P, Vihman MM (in press) Language acquisition and phonological change. In: Honeybone P, Salmons JC (eds) The handbook of historical phonology. OUP, Oxford

35. Foulkes P, Docherty G, Watt D (2005) Phonological variation in child directed speech. Language 81:177–206
36. Fowler C (1984) Segmentation of coarticulated speech in perception. Percept Psychophys 36:359–368
37. Fowler C, Brown J (2000) Perceptual parsing of acoustic consequences of velum lowering from information for vowels. Percept Psychophys 62:21–32
38. Fowler C, Housum J (1987) Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. J Mem Lang 26:489–504
39. Fowler C, Smith M (1986) 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. Erlbaum, Hillsdale, pp 123–139
40. Fowler C, Richardson M, Marsh K, Shockley K (2008) Language use, coordination, and the emergence of cooperative action. In: Fuchs A, Jirsa V (eds) Understanding complex systems. Springer, Berlin, pp 261–279
41. Fujisaki H, Kunisaki O (1976) Analysis, recognition and perception of voiceless fricative consonants in Japanese. Annu Bull Res Inst Logop Phoniatr 10:145–156
42. Garrett A, Johnson K (2013) Phonetic bias in sound change. In: Yu A (ed) Origins of sound change. Oxford University Press, Oxford, pp 51–97
43. Garrod S, Pickering M (2009) Joint action, interactive alignment, and dialog. Top Cogn Sci 1:292–304
44. Goldinger S (1997) Words and voices: perception and production in an episodic lexicon. In: Johnson K, Mullennix J (eds) Talker variability in speech processing. Academic Press, San Diego, pp 33–66
45. Grammont M (1902) Observations sur le langage des enfants. In Me langes Linguistiques Offerts aI M. Antoine Meillet. Klincksieck, Paris, pp 115–131
46. Greenlee M, Ohala J (1980) Phonetically motivated parallels between child phonology and historical sound change. Lang Sci 2:283–301
47. Grosvald M, Corina D (2012) 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. John Benjamins, Amsterdam, pp 77–100
48. Guion S (1998) The role of perception in the sound change of velar palatalization. Phonetica 55:18–52
49. Halle M, Hughes G, Radley J-P (1957) Acoustic properties of stop consonants. J Acoust Soc Am 29:107–116
50. Harrington J (2006) An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. J Phon 34:439–457
51. Harrington J (2010) Acoustic phonetics. In: Hardcastle W, Laver J, Gibbon F (eds) A handbook of phonetics. Wiley-Blackwell, Oxford, pp 81–129
52. Harrington J (2012) The relationship between synchronic variation and diachronic change. In: Cohn A, Fougeron C, Huffman M (eds) Handbook of laboratory phonology. Oxford University Press, Oxford, pp 321–332
53. Harrington J, Palethorpe S, Watson C (2000) Does the Queen speak the Queen's English? Nature 408:927–928
54. Harrington J, Kleber F, 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
55. Harrington J, Hoole P, Kleber F, Reubold U (2011) The physiological, acoustic, and perceptual basis of high back vowel fronting: evidence from German tense and lax vowels. J Phon 39:121–131
56. Harrington J, Kleber F, Reubold U (2012) The production and perception of coarticulation in two types of sound change in progress. In: Fuchs S, Weirich M, Perrier P, Pape D (eds) Speech production and speech perception: planning and dynamics. Peter Lang, Bern, pp 33–55

57. Harrington J, Kleber F, Reubold U (2013) The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. J Acoust Soc Am 134:551–561
58. Harrington J, Kleber F, Stevens M (2015) The relationship between the (mis)-parsing of coarticulation in perception and sound change: evidence from dissimilation and language acquisition. In: Esposito A, Faundez-Zany M (eds) Recent advances in nonlinear speech processing. Springer, Berlin
59. Harrington J, Kleber F, Reubold U, Siddins J (2015) The implications for sound change of prosodic weakening: evidence from polysyllabic shortening. Lab Phonol 6(1):87–117
60. Hawkins S (2003) Roles and representations of systematic fine phonetic detail in speech understanding. J Phon 31:373–405
61. Hay J, Foulkes P (in press) The evolution of medial /t/ over real and remembered time. Language
62. Hay J, Jannedy S, Mendoza-Denton N (1999) Oprah and /ay/: lexical frequency, referee design and style. In: Proceedings of the 14th international congress of phonetic sciences, San Francisco
63. Hay J, Pierrehumbert J, Walker A, LaShell P (2015) Tracking word frequency effects through 130 years of sound change. Cognition 139:83–91
64. Heid S, Hawkins S (2000) An acoustical study of long domain /r/ and /l/ coarticulation. In: Proceedings of the 5th seminar on speech production: models and data. Kloster Seeon, Bavaria, Germany. Munich, pp 77–80
65. Hintzman DL (1986) 'Schema abstraction' in a multiple-trace memory model. Psychol Rev 93:328–338
66. Hockett C (1950) Age-grading and linguistic continuity. Language 26:449–457
67. Hoole P, Pouplier M, Benus S, Bombien L (2013) Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: data and modelling. In: Spreafico L, Vietti A (eds) Rhotics: new data and perspectives. Bolzano University Press, pp 81–97
68. Hualde J (2011) Sound change. In: van Oostendorp M, Ewen C, Hume E, Rice K (eds) The blackwell companion to phonology, vol IV. Wiley-Blackwell, Oxford, pp 2214–2235
69. Hyman LM (2013) Enlarging the scope of phonologization. In: Yu A (ed) Origins of sound change: approaches to phonologization. Oxford University Press, Oxford, pp 3–28
70. Jakobson R (1941) Kindersprache, Aphasie, und allgemeine Lautgesetze. Almqvist & Wiksell, Uppsala
71. Janda R (2003) "Phonologization" as the start of dephoneticization—or, one sound change and its aftermath: of extension, generalization, lexicalization, and morphologization. In: Joseph B, Janda R (eds) The handbook of historical linguistics. Blackwell, Oxford, pp 401–422
72. Janda R, Joseph B (2003) Reconsidering the canons of sound-change: towards a "big bang" theory. In: Blake B, Burridge K (eds) Selected papers from the 15th international conference on historical linguistics. John Benjamins, Amsterdam, pp 205–219
73. Johnson K (1997) Speech perception without speaker normalization: An exemplar model. In: Johnson K, Mullennix J (eds) Talker variability in speech processing. Academic Press, San Diego, pp 145–165
74. Kataoka R (2011) Phonetic and cognitive bases of sound change. Ph.D diss. University of California, Berkeley
75. Kawasaki H (1986) Phonetic explanation for phonological universals: the case of distinctive vowel nasalization. In: Ohala J, Jaeger J (eds) Experimental phonology. Academic Press, Orlando, pp 81–103
76. Kerswill P (1996) Children, adolescents, and language change. Language Variation and Change 8:177–202
77. Kingston J, Diehl RL (1994) Phonetic knowledge. Language 70:419–454
78. Kirby J (2013) The role of probabilistic enhancement in phonologization. In: Yu A (ed) Origins of sound change: approaches to phonologization. Oxford University Press, Oxford, pp 228–246

79. Kirby J (2014) Incipient tonogenesis in Phnom Penh Khmer: computational studies. Laboratory phonology 5:195–230
80. Kiparsky P (2008) Universals constrain change; change results in typological generalizations. In: Good J (ed) Linguistic universals and language change. Oxford University Press, Oxford, pp 23–53
81. Kleber F, Peters S (2014) Children's imitation of coarticulatory patterns in different prosodic contexts. In: Proceedings of 14th conference on laboratory phonology, Tokyo, Japan
82. Kleber F, Harrington J, Reubold U (2012) The relationship between the perception and production of coarticulation during a sound change in progress. Lang Speech 55:383–405
83. Labov W (1981) Resolving the Neogrammarian controversy. Language 57:267–308
84. Labov W (1990) The intersection of sex and social class in the course of linguistic change. Lang Var Chang 2:205–254
85. Labov W (1994) Principles of linguistic change. Volume 1 internal factors. Blackwell, Malden
86. Labov W (2001) Principles of linguistic change. Volume 2: social factors. Blackwell, Oxford
87. Labov W (2007) Transmission and diffusion. Language 83:344–387
88. Labov W (2010) Principles of linguistic change, Volume 3: cognitive and cultural factors. Wiley-Blackwell, Oxford
89. Lehiste I (1970) Suprasegmentals. MIT Press, Cambridge
90. Lightfoot D (1999) The development of language: acquisition, change, and evolution. Blackwell, Malden
91. Lin S, Beddor P, Coetzee A (2014) Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. Lab Phonol 5:9–36
92. Lindblom B (1990) Explaining phonetic variation: a sketch of the H & H theory. In: Hardcastle W, Marchal A (eds) Speech production and speech modeling. Kluwer, Dordrecht, pp 403–439
93. Lindblom B, Studdert-Kennedy M (1967) On the role of formant transitions in vowel recognition. J Acoust Soc Am 42:830–843
94. Lindblom B, Guion S, Hura S, Moon S-J, Willerman R (1995) Is sound change adaptive? Rivista di Linguistica 7:5–36
95. Lobanov B (1971) Classification of Russian vowels spoken by different speakers. J Acoust Soc Am 49:606–608
96. Local J (2003) Variable domains and variable relevance: interpreting phonetic exponents. J Phon 31:321–329
97. Luick K (1964) Historische Grammatik der englischen Sprache. Blackwell, Oxford
98. Maclagan M, Hay J (2007) Getting fed up with our feet: contrast maintenance and the New Zealand English "short" front vowel shift. Lang Var Chang 9:1–25
99. Mann V, Repp B (1980) Influence of vocalic context on the perception of [ʃ]–[s] distinction: I. Temporal factors. Percept Psychophys 28:213–228
100. Martinet A (1955) Economie des Changements Phone tiques. Francke, Bern
101. Matthies M, Perrier P, Perkell JS, Zandipour M (2001) Variation in anticipatory coarticulation with changes in clarity and rate. J Speech Lang Hear Res 44:340–353
102. Mowrey R, Pagliuca W (1995) The reductive character of articulatory evolution. Rivista di Linguistica 7:37–124
103. Munson B, Beckman M, Edwards J (2012) Abstraction and specificity in early lexical representations: climbing the ladder of abstraction. In: Cohn A, Fougeron C, Huffman M (eds) The oxford handbook of laboratory phonology. Oxford University Press, Oxford, pp 288–309
104. Müller D (2013) Liquid dissimilation with a special regard to Latin. In: Sánchez Miret F, Recasens D (eds) Studies in phonetics, phonology and sound change in romance. Lincom, Munich, pp 95–109
105. Müller V, Harrington J, Kleber F, Reubold U (2011) Age-dependent differences in the neutralization of the intervocalic voicing contrast: evidence from an apparent-time study on East Franconian. Interspeech, Florence
106. Nielsen K (2011) Specificity and abstractness of VOT imitation. J Phon 39:132–142
107. Nielsen K (2014) Phonetic imitation by young children and its developmental changes. J Speech Lang Hear Res 57:2065–2075

108. Nittrouer S, Studdert-Kennedy M (1987) The role of coarticulatory effects in the perception of fricatives by children and adults. J Speech Hear Res 30:319–329
109. Nosofsky R (1988) Exemplar-based accounts of relations between classification, recognition, and typicality. J Exp Psychol Learn Mem Cogn 14:700–708
110. Öhman S (1966) Coarticulation in VCV utterances: spectrographic measurements. J Acoust Soc Am 39:151–168
111. Ohala J (1981) The listener as a source of sound change. In: Masek CS, Hendrick RA, Miller MF (eds) Papers from the parasession on language and behavior. Chicago Linguistic Society, Chicago, pp 178–203
112. Ohala J (1993) The phonetics of sound change. In: Jones C (ed) Historical linguistics: problems and perspectives. Longman, London, pp 237–278
113. Ohala J (2012) 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. John Benjamins, Amsterdam, pp 21–36
114. Ohala M, Ohala J (1992) Phonetic universals and Hindi segment duration. In: Proceedings of the international conference on spoken language processing. Edmonton, Alberta, pp 831–834
115. Osthoff H, Brugmann K (1878) Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen, Band I. Leipzig
116. Pardo J, Jay I, Krauss R (2010) Conversational role influences speech imitation. Atten Percept Psychophys 72:2254–2264
117. Pardo J, Gibbons R, Suppes A, Krauss R (2012) Phonetic convergence in college roommates. J Phon 40:190–197
118. Paul H (1886) Prinzipien der Sprachgeschichte, 2nd edn. Halle, Niemeyer
119. Phillips B (2006) Word frequency and lexical diffusion. Palgrave Macmillan, Basingstoke
120. Pierrehumbert J (2001) Exemplar dynamics: word frequency, lenition, and contrast. In: Bybee J, Hopper P (eds) Frequency effects and the emergence of lexical structure. John Benjamins, Amsterdam, pp 137–157
121. Pierrehumbert J (2003) Phonetic diversity, statistical learning, and acquisition of phonology. Lang Speech 46:115–154
122. Pierrehumbert J (2006) The next toolkit. J Phon 34:516–530
123. Pierrehumbert J, Beckman M, Ladd DR (2000) Conceptual foundations of phonology as a laboratory science. In: Burton-Roberts N, Carr P, Docherty G (eds) Phonological knowledge. Oxford University Press, Oxford, pp 273–303
124. Pierrehumbert J, Stonedahl F, Dalaud R (2014) A model of grassroots changes in linguistic systems. arXiv:1408.1985v1
125. Quené H (2013) Longitudinal trends in speech tempo: the case of Queen Beatrix. J Acoust Soc Am 133:452–457
126. Repp BH (1982) Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. Psychol Bull 92:81–110
127. Reubold U, Harrington J (In press) The influence of age on estimating sound change acoustically from longitudinal data. In: Wagner SE, Buchstaller I (eds) Panel studies of variation and change. Routledge Ltd, New York
128. Ruch H, Harrington J (2014) Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. J Phon 45:12–25
129. Sancier M, Fowler C (1997) Gestural drift in a bilingual speaker of Brazilian Portuguese and English. J Phon 25:421–436
130. Sankoff G, Blondeau H (2007) Language change across the lifespan: /r/ in Montreal Speech. Language 83:560–588
131. Schuchardt H (1885) Über die Lautgesetze: Gegen die Junggrammatiker. Oppenheim, Berlin
132. Sebanz N, Bekkering H, Knoblich G (2006) Joint action: bodies and minds moving together. Trends Cogn Sci 10:70–76
133. Shockley K, Richardson D, Dale R (2009) Conversation and coordinative structures. Top Cogn Sci 1:305–319

134. Siddins J, Harrington J, Reubold U, Kleber F (2014) Investigating the relationship between accentuation, vowel tensity and compensatory shortening. In: Proceedings of 7th speech prosody conference. Dublin, Ireland
135. Silverman D (2006) the diachrony of labiality in trique, and the functional relevance of gradience and variation. In: Goldstein L, Whalen D, Best C (eds) Papers in laboratory phonology 8. New Haven, Mouton de Gruyter, pp 133–154
136. Solé M (2007) Controlled and mechanical properites in speech: a review of the literature. In: Solé M-J, Beddor P, Ohala M (eds) Experimental approaches to phonology. Oxford University Press, Oxford, pp 302–321
137. Solé M (2009) Acoustic and aerodynamic factors in the interaction of features: the case of nasality and voicing. In: Vigário M, Frota S, Freitas MJ (eds) Phonetics and phonology: interactions and interrelations. John Benjamins, Amsterdam, pp 205–234
138. Solé M (2014) The perception of voice-initiating gestures. Lab Phonol 5:37–68
139. Stanford J, Kenny L (2013) Revisiting transmission and diffusion: an agent-based model of vowel chain shifts across large communities. Lang Var Chang 25:119–153
140. Stevens K (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In: David E, Denes P (eds) Human communication: a unified view. McGraw-Hill, New York, pp 51–66
141. Stevens K (1989) On the quantal nature of speech. J Phon 17:3–45
142. Stevens M, Harrington J (2014) The individual and the actuation of sound change. Loquens 1(1):e003. doi:10.3989/loquens.2014.003
143. Stevens M, Reubold U (2014) Pre-aspiration, quantity, and sound change. Lab Phonol 5:455–488
144. Studdert-Kennedy M (1998) Introduction: the emergence of phonology. In: Hurford J, Studdert- Kennedy M, Knight C (eds) Approaches to the evolution of language. Cambridge University Press, Cambridge, pp 169–176
145. Stampe D (1969) The acquisition of phonetic representation. In: Papers from the fifth regional meeting of the chicago linguistic society, pp 443–454
146. Sweet H (1888) A history of English sounds. Clarendon Press, Oxford
147. Torreira F (2012) Investigating the nature of aspirated stops in Western Andalusian Spanish. J Int Phon Assoc 42:49–63
148. Trudgill P (1974) Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. Lang Soc 3:215–246
149. Trudgill P (1983) On dialect: social and geographical perspectives. New York University Press, New York
150. Trudgill P (2004) Dialect contact and new-dialect formation: the inevitability of colonial englishes. Edinburgh University Press, Edinburgh
151. Trudgill P (2008) Colonial dialect contact in the history of European languages: on the irrelevance of identity to new-dialect formation. Lang Soc 37:241–254
152. Trudgill P (2008) On the role of children, and the mechanical view: a rejoinder. Lang Soc 37:277–280
153. Vihman M (1980) Sound change and child language. In: Traugott E, La Brum R, Shepherd S (eds) Papers from the 4th international conference on historical linguistics. John Benjamins, Amsterdam, pp 303–320
154. Wang WS-Y (1969) Competing changes as a cause of residue. Language 45:9–25
155. Wardhaugh R, Fuller J (2015) An introduction to sociolinguistics, 7th edn. Wiley Blackwell, Chichester
156. Watson CI, Maclagan M, Harrington J (2000) Acoustic evidence for vowel change in New Zealand English. Lang Var Chang 12:51–68
157. Wedel A (2007) Feedback and regularity in the lexion. Phonology 24:147–185
158. Weinreich U, Labov W, Herzog M (1968) Empirical foundations for a theory of language change. In: Lehmann W, Malkiel Y (eds) Directions for historical linguistics. University of Texas Press, Austin, pp 95–195
159. Wells J (1982) Accents of English 2: the British Isles. Cambridge University Press, Cambridge

160. Wolfram W, Schilling-Estes N (2003) Dialectology and diffusion. In: Joseph B, Janda R (eds) The handbook of historical linguistics. Blackwell, Oxford, pp 713–735
161. Wright R (2003) Factors of lexical competition in vowel articulation. In: Local J, Ogden R, Temple R (eds) Papers in laboratory phonology VI. Cambridge University Press, Cambridge, pp 75–87
162. Yu A (2013) Individual differences in socio-cognitive processing and the actuation of sound change. In: Yu A (ed) Origins of sound change: approaches to phonologization. Oxford University Press, Oxford, pp 201–227
163. Yu A, Abrego-Collier C, Sonderegger M (2013) Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. Plos One 8(9):e74746
164. Zellou G, Tamminga M (2014) Nasal coarticulation changes over time in Philadelphia English. J Phon 47:18–35

jmh@phonetik.uni-muenchen.de