Phonology **25** (2008) 217–270. © 2008 Cambridge University Press
doi:10.1017/S0952675708001474 Printed in the United Kingdom

*The evolution of auditory dispersion in bidirectional constraint grammars**

Paul Boersma
University of Amsterdam

Silke Hamann
University of Düsseldorf

This paper reconciles the standpoint that language users do not aim at improving their sound systems with the observation that languages seem to improve their sound systems. If learners optimise their perception by gradually ranking their cue constraints, and reuse the resulting ranking in production, they automatically introduce a PROTOTYPE EFFECT, which can be counteracted by an ARTICULATORY EFFECT. If the two effects are of unequal size, the learner will end up with a sound system auditorily different from that of her language environment. Computer simulations of sibilant inventories show that, independently of the initial auditory sound system, a stable equilibrium is reached within a small number of generations. In this stable state, the dispersion of the sibilants of the language strikes an optimal balance between articulatory ease and auditory contrast. Crucially, these results are derived within a model without any goal-oriented elements such as dispersion constraints.

1 Introduction

It has often been observed that sound systems are structured in a way that minimises the perceptual confusion between its elements. For instance, a language with only three vowels tends to keep them far apart in the two-dimensional space of the auditory first and second formants, i.e. such a language tends to pronounce them as sounds close to the 'corner' vowels [a i u]. A related observation is the existence of chain shifts in sound change. For instance, a change of a vowel pronounced [u] into a vowel

* The ideas and simulations in this paper were presented at OCP 3 in Budapest in January 2006, and at the 29th Annual Meeting of the DGfS in Siegen in February 2007. We thank Jaye Padgett and the audiences at these talks, especially Laura Downing and Andy Wedel, for useful discussion. This research was supported by grants 016.024.018 and 016.064.057 from the Netherlands Organisation for Scientific Research (NWO).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

pronounced [y] is often followed by a change of [o] into [u], filling up the corner just vacated; or the consonants pronounced [dʃ d t] in Proto-Indo-European have shifted to [d t θ] in Germanic and on to [t ts d] in German. In all these shifts, the common theme is that the contrast between the members of the inventory is maintained, although all members change. At the abstract level of the language, therefore, it appears that languages actively strive to implement an OPTIMAL AUDITORY DISPERSION and to maintain this dispersion diachronically.

At the same time, many authors insist that perceptually based sound change can only take place by INNOCENT MISAPPREHENSION, i.e. that speakers do not have the perceptual optimisation of a sound system as a goal but that sound change is instead caused by learners who reanalyse the imperfectly transmitted sounds of their language environment (Ohala 1981, Blevins 2004).

There seems to be a tension between the idea of optimal auditory dispersion and the idea of innocent misapprehension. A proponent of auditory dispersion even claims that 'sound change through misperception … can only hope to account for neutralization, not dispersion or enhancement' (Flemming 2005: 173). This is not entirely true: there may exist non-goal-oriented mechanisms by which improvement of auditory contrast could be a common but unintended *result* of innocent misapprehension. For instance, Blevins (2004: 285–289) tentatively explains chain shifts with the help of Pierrehumbert's (2001) claim that exemplar theory predicts automatic shifting of auditory vowel prototypes to regions where they are less likely to be perceived as a different category (see §7.2). All authors agree, however, that formalising auditory dispersion with existing formal phonological devices such as Optimality Theory (OT) is incompatible with the claim of innocent misapprehension. Thus, the OT accounts of dispersion of Flemming (1995, 2004), Padgett (2001, 2003a, b, 2004) and Sanders (2003) contain explicitly goal-oriented elements, namely dispersion constraints.

The present paper reconciles auditory dispersion with innocent misapprehension without resorting to exemplar theory. We show that in the other main neurologically informed linguistic framework, namely constraint grammars (stochastic OT, noisy Harmonic Grammar, Maximum Entropy), auditory dispersion does turn out to emerge mechanically as long as we assume bidirectionality, i.e. that a language user applies the same constraint rankings (or weights) to perception and production. We thus show that the innocent misapprehension standpoint can be correct in stating that speakers are not goal-oriented, while at the same time the auditory dispersion standpoint can be correct in observing that sound change tends to minimise perceptual confusion. The reconciliation will be seen to derive from the possibility that sound change is teleological at the abstract level of the observed language, but non-teleological at the concrete level of the language user; this situation is analogous to that in evolutionary biology (Darwin 1859), where adaptations to the environment are observationally optimising but underlyingly non-teleological.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

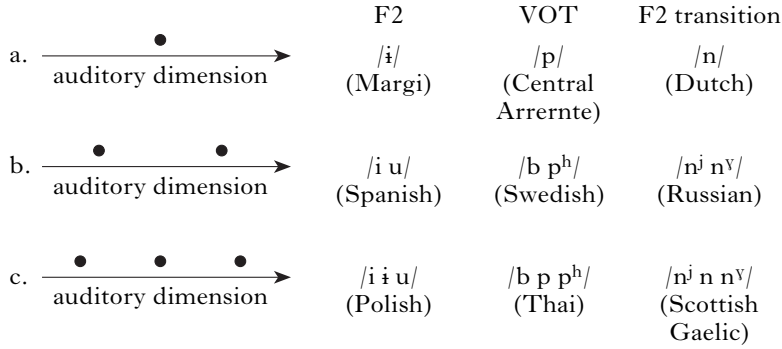


Figure 1
Observed auditory dispersion.

This paper is structured as follows. In §2 we provide an overview of dispersion effects, and briefly discuss earlier accounts of them. In §3 we introduce the bidirectional grammar model that we need for our own non-teleological explanation of dispersion. In §4 we introduce the specific case considered in this article, sibilant dispersion. In §§5 and 6 we provide computer simulations of the acquisition of sibilant inventories, which show that non-dispersed inventories can be learned to some extent but will automatically change within a few generations into dispersed systems that will stay stable over any following generations. In §7 we discuss the necessary assumptions in our model and compare them with those of previous models. We end by concluding that dispersion effects are intrinsic to bidirectional constraint-based grammar models, so that we predict analogous effects operating at all levels of the grammar.

2 Auditory dispersion effects and explanations

In this paper we confine ourselves to the simplest case of dispersion, namely the one-dimensional case. In this section we first present six kinds of dispersion effects, then discuss earlier accounts from the non-OT and OT literature.

2.1 Six kinds of auditory dispersion effects

In Fig. 1 we see examples of how phonological categories can be dispersed along one-dimensional auditory continua. This figure helps us to illustrate the six kinds of dispersion effects. Most languages tend to be dispersed like the cases we discuss here; cases of real or apparent undispersed inventories are discussed at the end of this section.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

2.1.1 *The preference for the centre.* If a language has only one category on the continuum, it tends to be in the centre (Fig. 1a). We know this because it tends to have the same auditory value as the mid value of an inventory with three phonemes on the continuum (Fig. 1c). For instance, languages with only one category on the continuum of voice onset time (VOT) for plosives, such as Central Arrernte (Breen & Dobson 2005), typically have only the zero VOT value, as in the plain voiceless plosive [p],[1] while languages with three categories, such as Thai (Tingsabadh & Abramson 1999), typically have a 'prevoiced' [b] (negative VOT), [p] (zero VOT) and an aspirated [p^h] (positive VOT). Likewise, languages with only one high vowel, such as Margi (Maddieson 1987) or Kabardian (Choi 1991), tend to have a vowel with a mid second formant (F2) value, such as [ɨ], while languages with three high vowels, such as Polish (Jassem 2003), most often have a rounded back vowel [u] (low F2), a central vowel [ɨ] (mid F2) and a spread front vowel [i] (high F2). Finally, languages with only one value on the palatalised–velarised continuum for alveolar nasals, such as English, tend to have the plain [n] (mid F2 transition), while languages with three such nasals, such as Scottish Gaelic (Borgstrøm 1940), tend to have a palatalised [n^j] (high F2 transition), a plain [n] and a velarised [$n^ɣ$] (low F2 transition).[2] In all these cases, the single value of the inventories with one category equals the mid value of the inventories with three categories. For the tendency of single-category and multiple-category inventories to share a category, two explanations have been proposed. The explanation of FEATURAL MARKEDNESS (Jakobson 1941) claims that the shared category ([p] or [n]) reflects a phoneme with UNMARKED feature values (/p/ = [−voiced, −spread glottis], /n/ = [−front, −back]), whereas the other categories reflect MARKED feature values (/b/ = [+voiced], /p^h/ = [+spread glottis], /n^j/ = [+front], /$n^ɣ$/ = [+back]). And the explanation of ARTICULATORY EFFORT (Lindblom 1990b) claims that the shared category is articulatorily easiest or BASIC (thus [p] = no or little laryngeal activity,[3] [n] = no special tongue body movements), whereas the other categories often involve additional gestures and can be called ELABORATE ([b] = active vocal fold adduction, [p^h] = large active vocal fold abduction, [n^j] = tongue body fronting, [$n^ɣ$] = tongue body backing). In the latter view, the fact that the shared category tends to have the auditory *mid* value could be explained by the consideration that one value must be the easiest, and that humans have a large variety of articulatory tricks at their disposal, so that they are capable of producing gestures that deviate from this easiest

[1] We use the symbol [p] to cover both the lenis voiceless [b̥] and any more fortis varieties. For most languages the sources do not make a principled distinction.

[2] In Bernera Gaelic, Ladefoged *et al.* (1997) did not find the three-way contrast in nasals, but they did find it in laterals.

[3] Whether it is 'no' or 'little' depends on the amount of active devoicing, i.e. the activity of the posterior cricoarytenoid muscle. Some languages have no posterior cricoarytenoid activity during [b̥/p] (e.g. English: Hirose & Gay 1972); some languages do have such activity, but it is shorter and less strong than in [p^h] (e.g. Danish: Hutters 1985).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

gesture in opposite directions, with presumably opposite auditory results. This explanation for the centrality of the single category does seem to hold for the examples just discussed, and the idea that articulations resulting in peripheral auditory values involve more effort than those resulting in more central values has been used in computational models of inventories, such as the vowel model of ten Bosch (1991). A third view, which downplays the role of articulatory effort and instead stresses auditory distinctiveness, takes into account inventories with two categories on the continuum. This we describe next.

2.1.2 *The excluded centre.* Languages with two categories on the continuum often have them on two sides of the centre. With Flemming (1995) and Padgett (2001, 2003a, b), we regard this phenomenon as the crucial piece of evidence for the existence of auditory dispersion as a driving force for inventories. Flemming (1995, 2004, 2005, 2006), for instance, notes that a language with two high vowels tends to have [i] (high F2) and [u] (low F2), crucially excluding the intermediate [ɨ] vowel, which is the one that languages with one high vowel such as Margi and Kabardian tend to have. He then argues that the usual account of inventories in terms of markedness of phonemes does not work: since [i] and [u] are much more common cross-linguistically than [ɨ], and [i] slightly more common than [u], a markedness account would probably regard /i/ as the unmarked high vowel and would therefore predict that languages with a single high vowel have /i/. Padgett makes a similar argument for the palatalisation–velarisation continuum in Russian, which has [nʲ] and [nˠ] but not the arguably unmarked and least effortful [n]. The only explanation is that in the [i]–[u] and [nʲ]–[nˠ] pairs, the concept of auditory contrast plays a decisive role. With featural markedness out of the game, the explanation for the [p], [ɨ] and [n] singletons discussed in the previous paragraph must be articulatory. But Lindblom's basic–elaborated opposition does not do the job either: the easiest high vowel must be [ɨ] (it has the smallest articulatory distance from the neutral tongue shape [ə]), and indeed it appears as the only vowel in languages with one high vowel; but the 'basic' [ɨ] does not occur in languages with two high vowels, and this strongly suggests that the larger auditory confusability within an inventory like [i ɨ] plays a stronger role than the larger articulatory effort associated with the inventory [i u]. This kind of comparison between inventories is precisely what ten Bosch, Flemming and Padgett have tried to capture in their models, which formalise the competition between articulatory and auditory considerations.

2.1.3 *Equal auditory distances.* The effects mentioned in §§2.1.1 and 2.1.2 can be summarised as a single PRIMARY AUDITORY DISPERSION EFFECT: categories seem to be located within the auditory space in such a way that they are perceptually maximally distinct. For languages with three categories along a continuum, this idea predicts that they should typically have the middle category spaced at equal auditory distances from

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

its neighbours. We can check this when looking at languages with four categories along the continuum. For instance, an optimal dispersion of the auditory vowel height continuum (first formant) involves the middle value of the triplet [i 'e' a] (e.g. Bradlow 1995 for Spanish) lying between the two mid values of the quadruplet [i e ɛ a] (e.g. Harrison 1997 for Catalan).

2.1.4 *The growing space.* A SECONDARY AUDITORY DISPERSION EFFECT is that for larger inventories, the auditory space enlarges, but the distance between the categories decreases. In Fig. 1, for instance, the auditory space taken up by the inventory with three elements is larger than the auditory space that has to accommodate only two elements. For instance, the high back vowel is often auditorily fronted in languages with two high vowels (e.g. English [ʉː], Japanese [ɯ]), but not in languages with three high vowels. In languages with many vowels, the three corners of the vowel space tend to be [a i u], whereas languages with only three vowels (such as Tagalog) often have the reduced [ɐ ɪ ʊ] (for an overview, see Boersma 1998: 216); and languages with no more than a single vowel (such as several Germanic languages in unstressed syllables) typically have just [ə] (Flemming 2004: 235, 2005: 164). A plausible explanation is that the more peripheral a value is, the more articulatory effort tends to be required to implement it; languages with two categories require a smaller total auditory space for obtaining a small perceptual confusability than languages with three categories do, so that the balance between auditory contrast and articulatory effort turns out differently for the two cases (note that this is a goal-oriented description at the level of the language, not implying that there is any goal-orientedness in the underlying mechanism).

2.1.5 *Permissible variation.* If the phenomena described in the previous paragraphs really have to do with auditory contrast, this would predict that a category is allowed more auditory variation if it is alone on its auditory continuum than if it has neighbours from which it has to stay distinct. This is borne out by the data. Languages with a single labial plosive [p] typically have voiced allophones such as [b] in postnasal or intervocalic position (e.g. Central Arrernte: Breen & Dobson 2005), and languages with a single high vowel [ɨ] typically have allophones everywhere between [i] and [u], depending on the surrounding consonants (e.g. Kabardian: Choi 1991). Likewise, languages with smaller vowel inventories show more vowel-to-vowel coarticulation (e.g. Manuel 1990 for Bantu languages). There is no doubt that syntagmatic articulatory optimisation is involved. Again we see an interplay between the demands of auditory contrast and the demands of articulatory economy.

2.1.6 *Chain shifts.* The five effects just mentioned are all STATIC effects: they can be seen in synchronic inventories. There also exist two kinds of DYNAMIC effects, which can be seen in the diachronic development

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

of inventories: in a PUSH CHAIN, one category approaches another and seemingly pushes it away, and in a DRAG CHAIN, one category vacates a region on the auditory continuum, thereby seemingly allowing another category to fill up the vacated space. In order to account for these chain shifts, a model has to exhibit properties that lead to a repulsive force between categories. Not all explanations of inventories have this property. In some theories of inventories a restricted kind of dispersion comes about by selective neutralisation of categories (categories that are auditorily close to each other have a greater chance of merging into a single category, so that the remaining categories tend to be spaced further apart; e.g. Kochetov 2008) or by unsupervised clustering of proto-categories (e.g. de Boer 1999, Oudeyer 2006). Such theories, in which close categories attract rather than repel one another, can only account for diachronic merger, not for chain shifts, and must therefore be left out of consideration in the present paper, because even the strongest proponents of innocent misapprehension agree that 'there is no question that chain shifts exist' (Blevins 2004: 285).

It must be clearly stated here that all these effects are tendencies rather than fixed rules. For one thing, the language may be in a transitory, non-equilibrium state, and this may involve non-optimal auditory dispersion (as will become clear in our simulations in §§5 and 6). Next, considerations of articulatory effort may limit auditory contrast to the extent that an inventory with two categories can include the centre value, so that it is asymmetric along the auditory continuum. For instance, while the optimally dispersed [b pʰ] inventory does exist in Swedish (Jakobson 1941, Ringen & Helgason 2004) and several other languages (Keating *et al.* 1983), it does involve two separate articulatory gestures and is therefore often replaced with just a voicing contrast ([b p]: Dutch, French) or just an aspiration contrast ([p pʰ]: English, Mandarin), which are articulatorily easier (both contain the 'basic' category) and still maintain a sufficient auditory distinction (Lindblom 1990b calls this effect 'adaptive dispersion').[4] Finally, these single auditory continua live in an inventory with multiple continua, some of which may intrude. For instance, Flemming (1995: 31) explains the fact that in many languages /i u/ is realised as [i ʉ] or [i ɯ] as an enhancement of the /u/–/o/ distinction. Rather than contradicting the auditory dispersion idea, all these exceptions corroborate it by leading to explanations that involve articulatory effort, wider-scoped data and contingent histories, beside auditory dispersion, but never featural markedness.

The following two sections discuss various ways in which auditory dispersion has been modelled explicitly.

[4] It is also possible that the *acoustic* VOT continuum corresponds not to a single auditory VOT continuum, but to two separate auditory continua, such as periodic murmur and noisiness.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

2.2 Non-OT accounts of auditory dispersion

Auditory dispersion in vowel inventories was first modelled explicitly by Liljencrants & Lindblom (1972), who showed that two-dimensional vowel spaces (auditory height and backness) with a minimum probability of perceptual confusion look like real attested vowel inventories (some problems of detail were addressed by Lindblom 1986 and Schwartz *et al.* 1997).

Ten Bosch (1991) compared several techniques for optimising the distances between vowels, and found that the strategy of MAXIMISING THE MINIMAL DISTANCE worked best. If we start out with a random set of vowels and then iteratively move apart the two vowels that are closest to each other, we ultimately obtain an evenly dispersed inventory.

A possible criticism of this work is that the resulting vowel inventories are not symmetric enough. For instance, real vowel inventories tend to be structured in such a way that back vowels tend to have the same height contour as front vowels, yet there is no force in these models according to which such symmetries are enforced, and indeed the vowel inventories that result from the simulations of these authors tend to be asymmetric (Boersma 1998: 357). To remedy this problem, the simulations would have to be extended with tricks to incorporate an efficient use of available auditory, articulatory and/or phonological features.

A problem that is much more difficult to remedy is the inherent teleology in these models. In order to arrive at an optimal vowel inventory, a model typically starts with a random non-optimal vowel inventory, and tries to move towards an optimal end result by making small changes to the locations of the vowels in the vowel space, where every change has to be optimising, i.e. every change has to improve the auditory distinctiveness of the whole system. Even at the most concrete level of modelling, then, these models work with teleological devices.

Non-teleological accounts of auditory dispersion have been proposed as well. Blevins (2004: 285–289) sketches how within a framework where listeners store auditory events as exemplars in episodic memory and subsequently reuse these exemplars in production, following Pierrehumbert (2001), speakers may tend to choose exemplars that are little likely to be perceived as anything but the intended category. An explicit account of the details of such a scheme is provided by Wedel (2004, 2006). We discuss this in detail in §7.2.

Few of the non-OT accounts mentioned in this section make contact with phonological theory by modelling the interaction of the simulated inventories with phonological rules or constraints. Only exemplar models are likely to provide such a link: it seems perfectly possible to combine Wedel's (2006) dispersion model with his lexical analogy model (Wedel 2007), which, for example, implements violable metrical constraints as potentially conflicting biases that lead to the emergence of Gordon's (1999) frequency-informed regular patterns of weight-stress interaction. In the next section we discuss attempts to integrate the dispersion idea

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

into a tried and tested framework for phonological theory, namely Optimality Theory.

2.3 OT accounts of auditory dispersion

The original proposal of Optimality Theory by Prince & Smolensky (1993) handled inventories by constraint interaction and the device of 'richness of the base'. As in the featural markedness accounts mentioned above, a marked phonological element was only allowed to surface in a language if the unmarked counterpart of that phonological element also surfaced in that language. Prince & Smolensky's approach is therefore problematic for the same reason as mentioned earlier: it cannot account for the 'excluded centre' effect, where marked segments appear without their unmarked counterparts, as noted by Flemming (1995: 37, 2004: 235, 2005: 164, 2006: 250). So an OT account requires something more than just markedness and faithfulness constraints.

Flemming (1995) translated Lindblom's dispersion idea, and specifically ten Bosch's idea of maximising the minimum distance, into OT by introducing MINDIST constraints, which explicitly militate against inventories with small auditory distances between its members. Dispersion constraints such as Flemming's MINDIST, as well as the reformulations by Padgett (2001) and Sanders (2003), work very well in formalising the dispersion idea. However, they are explicitly teleological with respect to auditory dispersion. Furthermore, an empirical problem is that these constraints evaluate multiple inputs at a time: they can be said to evaluate whole inventories (Flemming 1995: 33–35, Boersma 1998: 361, McCarthy 2002: 226–227) or even entire languages (Padgett 2003a: 311, Flemming 2004: 268). These constraints are therefore hard to reconcile with the single-input constraints introduced by Prince & Smolensky, and the tableaux are hard to reconcile with tableaux that basically evaluate the processing of a single form in production (Prince & Smolensky 1993) or comprehension (Smolensky 1996). Flemming's and Padgett's general defence is that phonological theory is about possible languages, rather than about processing single forms. We feel that this standpoint underestimates the power of Optimality Theory as a decision mechanism: when used to evaluate single inputs, OT can be and has been applied successfully to processes such as production, comprehension and acquisition. If dispersion effects can be shown to emerge in OT even from modelling single-form processing, OT will not have to be invoked separately to evaluate the entire language.

The fact that Flemming, Padgett and Sanders handle dispersion effects by dedicated inherently teleological means (the dispersion constraints) in a synchronic grammar is sometimes seen as unproblematic (e.g. Hayes & Steriade 2004: 27), but a theory in which these effects arise automatically as side effects of more general independently needed devices should be preferred by Occam's razor, if such a theory exists. Padgett (2003b: 80) realises this shortcoming and suggests that dispersion constraints may just

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

express abstract observations about inventories while at the same time the real underlying mechanism may be more concrete and perhaps not explicitly goal-oriented. We agree, and in the present paper we provide just such an underlying mechanism within OT (in §7.6 we show that it also works in other constraint-based frameworks). Providing the underlying mechanism is necessary because we seek the locus of explanation in an acquisition bias on the part of the learner; formalising acquisition has to be done with the constraints that are in the learner's brain, not with constraints that describe behaviour at a higher level of abstraction.

In the present paper, we employ neither dispersion nor faithfulness constraints to account for dispersion effects. We show that dispersion effects can instead arise within a number of generations as the automatic result of CUE CONSTRAINTS, which are independently needed to model language-specific perception (e.g. Escudero & Boersma 2004), and ARTICULATORY CONSTRAINTS, which are independently needed to model articulatory effort in phonetic implementation (e.g. Boersma 1998, Kirchner 1998). The only assumption that we need to add to the pre-existing work on OT phonetics is that the speaker and the listener use the same grammar. The point is that the same constraints are used BIDIRECTIONALLY, i.e. both by the listener in comprehension and by the speaker in production. The following section illustrates this in more detail.

3 Bidirectional phonetics

A formal account of a linguistic phenomenon has to start by stating the representations involved. For our purposes we need only two, namely the (abstract, discrete) phonological SURFACE FORM and a (concrete, continuous) auditory-articulatory PHONETIC FORM. These two representations (see Fig. 2) are part of a more elaborate comprehensive model for bidirectional phonology and phonetics (Boersma 2007a), but any representations 'above' the surface form, such as the underlying form, are not discussed in this paper (except briefly in §§5.3 and 7.1), because they are not required for illustrating our point. Also, Fig. 2 keeps implicit any distinction between the auditory and articulatory parts of the phonetic form.

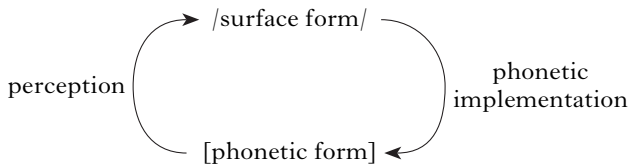


Figure 2
Processing models for phonetics and its interface with phonology.

The grammar model of Fig. 2 is bidirectional, i.e. it is used in two directions of processing: comprehension and production, as indicated by the

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

arrows in the figure. In the comprehension direction, the ('prelexical') PERCEPTION process maps an auditory-phonetic form to a phonological surface form; in the production direction, the PHONETIC IMPLEMENTATION process maps a phonological surface structure to an auditory-articulatory phonetic form.

We formalise the relations between the representations in Fig. 2 with violable constraints, as in Fig. 3.

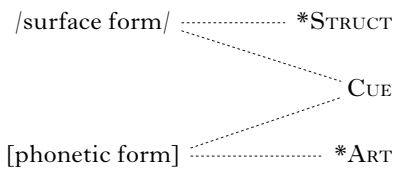


Figure 3
Grammar model for phonetics.

One of the reasons for modelling phonetics with constraints is that the output of the perception process tends to be restricted by the same structural constraints (*STRUCT in the figure) that have been proposed for phonological production. For instance, Polivanov (1931) proposes that Japanese learners of Russian perceive the Russian phonetic form [tak] as /.ta.ku./ because of a Japanese constraint on coda consonants, and the phonetic form [drama] as /.do.ra.ma./ because of a Japanese constraint against complex onsets.[5] Thus, the process that maps a phonetic form (or 'overt form' in the terms of Tesar & Smolensky 2000) to a phonological surface form is best regarded as being linguistic itself and therefore amenable to constraint-based modelling, a point previously made by Tesar (1997), Boersma (1998, 2007b), Tesar & Smolensky (1998, 2000) and Pater (2004).

We should point out here that modelling phonetic processing within a constraint-based framework such as OT does not imply that we regard low-level auditory and articulatory processing as belonging to a formal symbolic system specific to the human language faculty. On the contrary, a constraint-based decision mechanism like OT and/or Harmonic Grammar (HG; Smolensky & Legendre 2006) may well be typical of neural processing in general (Boersma 2003b: 444–445), and the success of OT in phonological theory may well be based on the fact that human phonological processing just uses this more general mechanism.

Having made plausible the claim that phonetic processing can be modelled with OT or HG, we can turn to the constraints proposed in Fig. 3. The relation between the phonological and the phonetic form is

[5] Polivanov's proposal was confirmed in perception experiments by Dupoux *et al.* (1999) and in brain activity experiments by Jacquemot *et al.* (2003). A reformulation in OT terms appeared in Escudero & Boersma (2004) and more explicitly in Boersma (2007b).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

evaluated by cue constraints (CUE), and the phonetic form on its own is evaluated by articulatory constraints (*ART). As an example of a cue constraint, consider the fact that the duration of a vowel is a major cue to the voicing of a following obstruent in English, both in perception (Denes 1955, Raphael 1972) and in production (Heffner 1937, House & Fairbanks 1953, Luce & Charles-Luce 1985), but at most a weak cue in some other languages, both in perception (Morrison 2002, Broersma 2005)[6] and in production (Keating 1979, 1985). Hence, the cue constraint *[long vowel duration] /obs,−voice/ is ranked high in English but low elsewhere.

Because of the perception–production symmetry just noted we follow Boersma (2006, 2007a, b) in assuming that the constraints in Fig. 3 are used bidirectionally. Bidirectionality in OT was previously proposed for faithfulness constraints by Smolensky (1996), for structural constraints by Tesar (1997) and Tesar & Smolensky (1998, 2000) and for cue constraints by Boersma (1998, in a control-loop model that does not show dispersion).

For Fig. 3, bidirectionality works as follows. In perception, the choice between candidate surface forms involves structural and cue constraints. In this direction of processing, the cue constraints evaluate language-specific cue integration (Escudero & Boersma 2004). For instance, the high ranking of *[long vowel duration] /obs,−voice/ in English predicts that an English listener, when confronted with an auditorily lengthened vowel, will be unlikely to perceive the following consonant as a voiceless obstruent, unless potentially conflicting constraints for other cues (or perhaps competing structural constraints) force her to. In phonetic implementation, the choice between candidate phonetic forms involves cue constraints and articulatory constraints. For instance, the high ranking of *[long vowel duration] /obs,−voice/ in English predicts that an English speaker, when intending to realise a voiceless obstruent, will be unlikely to lengthen the preceding vowel, unless articulatory constraints force her to.

We thus assume that speaker-listeners use cue constraints both in perception and in phonetic implementation, and that they rank (or weigh) these constraints identically in both directions of processing. The present paper shows that this bidirectional use of cue constraints leads to two asymmetries between perception and production, namely the PROTOTYPE EFFECT and the ARTICULATORY EFFECT, and that languages that are stable over the generations have to cancel these two biases out against one another, thus striking an optimal balance between minimisation of articulatory effort and minimisation of perceptual confusion, without there being any goal-oriented dispersion mechanism in the whole system.

[6] As a reviewer notes, these two references provide only indirect perceptual evidence, namely from the problems that L2 learners of English have acquiring this contrast. Direct proof of the existence of cross-dialectal differential cue weighting of a different contrast was found by Escudero & Boersma (2003, 2004).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

4 Sibilant inventories: dispersion of the spectral mean

The sounds with which we will exemplify the evolution of auditory contrast in this paper are the sibilants. We first make it plausible that sibilants indeed form a case of auditory dispersion.

The sibilants in a language can often be ordered along a one-dimensional acoustic continuum, namely the SPECTRAL CENTRE OF GRAVITY or SPECTRAL MEAN (e.g. Forrest *et al*. 1988, Gordon *et al*. 2002), as in Fig. 4.[7] Articulatorily, the spectral mean correlates with frontness of the tongue and with frontness of the place of articulation.

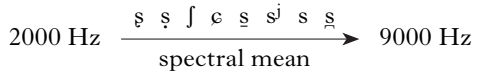


Figure 4
Scale of spectral mean for eight sibilants.

The acoustic spectral mean corresponds monotonically to an auditory continuum: when listening to a continuously interpolated [ʂ̻ʂ̺ʃɕs̻sʲsʂ̪] we hear *something* going up. This is not surprising: in the inner ear, as the sound [ʂ̻ʂ̺ʃɕs̻sʲsʂ̪] is unfolding, the region of excitation is moving along the basilar membrane, from the centre towards the oval window. We therefore assume that the auditory continuum involved can be described as an auditory spectral mean expressed along the Erb scale ('Equivalent Rectangular Bandwidth'; Moore & Glasberg 1983), which in the region of the sibilants correlates approximately logarithmically with the acoustic frequency in Hertz; for instance, the auditory distance from 2000 to 4000 Hz approximately equals the auditory distance from 4000 to 8000 Hz (they are 5·7 and 5·2 Erb respectively). We pretend in this paper that the auditory spectral mean is the only relevant perceptual cue for sibilants. In reality, other cues for sibilant articulator and place include spectral peaks (e.g. Jongman *et al*. 2000) and vowel transitions (e.g. Nowak 2006).

Sibilant inventories tend to be dispersed along the auditory spectral mean dimension, exhibiting all the effects mentioned in §2.1.

[7] Values reported in the literature can vary widely, which is partly due to the measurement method involved. In this paper, we assume that the spectral mean is computed by weighing the frequencies in the spectrum by their power densities (Forrest *et al*. 1988, Jongman *et al*. 2000). This is also the standard setting in Praat (Boersma & Weenink 1992–2008), which was used by Zygis & Hamann (2003), and by Padgett & Zygis (2003) in their subsequent OT modelling. The method yields values for sibilants that can indeed be as low as 2000 Hz or as high as 8000 Hz or more. As can be seen from inspecting their spectra and reported spectral means, Gordon *et al*. (2002) apparently used the incorrect method of Ladefoged (2003), which weighs the frequencies by their intensity values in dB and is therefore sensitive to arbitrary recording settings; this method tends to yield values very close to the centre of the frequency range, i.e. in Gordon *et al*.'s case very close to 5000 Hz.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Languages with only one sibilant, such as Spanish or Dutch (if we disregard the marginal and unstable Dutch alveopalatal; but see note 15) usually employ a sound with a fairly central spectral mean (variously classified as /s/ or /ʃ/ by English listeners) such as the Dutch flat laminal alveolar [s̻] (Mees & Collins 1982: 6) or the Spanish concave retracted apical alveolar [s̺] (Navarro Tomás 1932: 105–107, Harris 1969: 192). This is illustrated in Fig. 5a. Note that a single symbol can be employed for two widely different articulations, a sign that sibilants can be identified by their sound rather than by their articulation. The single sibilant is never at an edge of the spectral mean scale; although Maddieson (1984: 423) lists the retroflex [ʂ] as the only sibilant in the Dravidian language Kota, Flemming (2003: 354) convincingly illustrates that this is not the case: Kota [ʂ] is an allophone of the sibilant [s ~ ʧ] and occurs only adjacent to other retroflex sounds (Emeneau 1944).

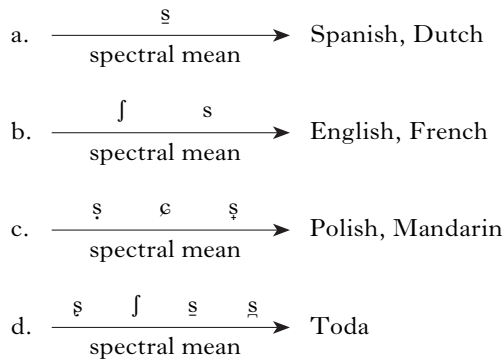


Figure 5
Dispersion of the spectral mean for inventories with one, two, three and four sibilants.

If a language has two sibilants, neither of them has a central spectral mean. Both sibilants are rather peripheral on the spectral mean dimension, as in English, which has a laminal, shallow-grooved and often rounded postalveolar [ʃ] and a deep-grooved alveolar [s] (Stone *et al.* 1992: 260). Again, the articulations can vary markedly between languages (for English and French: Dart 1991) and among speakers of the same language (Dart 1991), which is often due to anatomical differences (Toda 2005), but the auditory regions are similar.

In languages with a three-sibilant inventory, the three sibilant categories are equally spaced along the spectral mean dimension, usually implemented as (denti-)alveolar, alveopalatal and apical postalveolar, as in Polish (Puppel *et al.* 1977) and Mandarin (Ladefoged & Wu 1984); for an overview of the various possibilities for the 'apical postalveolar' see Hamann (2003). The sibilants produced by Polish women investigated by Zygis & Hamann (2003) have average acoustic spectral means of 3040,

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

4653 and 7439 Hz, which correspond to the evenly spaced auditory values of 24·40, 27·82 and 31·29 Erb. The three sibilants occupy a larger space than a two-sibilant inventory, illustrating the secondary dispersion effect (compare Fig. 5c with Fig. 5b). The Polish inventory is explicitly described in terms of auditory dispersion by Jones (2001), Padgett & Zygis (2003) and Zygis (2003).

Finally, languages with four sibilants, such as Toda, occupy an even larger space along the dimension of spectral mean: at the low end, Toda has a subapical palatal sibilant [s̻], which is the articulation with the lowest spectral mean of the attested sibilants. The descriptions for the other three Toda sibilants vary widely (Shalev *et al*. 1993, Ladefoged & Maddieson 1996: 156–160, Ladefoged 2001: 153), so not much more can be said.

The observation that all of the dispersion effects of §2.1 seem to be attested in sibilant inventories suggests two things. The first suggestion is that the spectral mean continuum indeed represents the main auditory cue for sibilant place; this makes it plausible that modelling sibilants along a single auditory continuum, which we do in §§5 and 6, is a valid approach. Independent evidence for this is that some perception experiments have been able to model the [s]–[ʃ] continuum by shifting a broad spectral plateau of constant width (in log Hz) along the frequency scale (Repp 1981). The second suggestion is that the articulatory-effort relations between the sibilants are similar to those of the auditory continua discussed in §2.1, i.e. that sibilants with central spectral mean values are easier to produce than those with peripheral values; this makes it plausible that including this articulatory-effort hypothesis in our modelling is a valid approach. Independent evidence for this comes from comparing the muscle activities required for producing central and peripheral sibilants. Sibilants with central spectral mean values like the Dutch flat laminal alveolar [s̻] or the Spanish apical alveolar [s̺] are produced by a simple raising of the articulators towards the roof of the mouth, without displacement or grooving of the tongue, and are thus as close as sibilants can get to the rest position of the tongue or to the average position of the tongue during vowels. At the high periphery of the spectrum, the deep-grooved alveolar sibilant (as in English) has a similar position to the Dutch non-grooved sibilant but requires additional activity of the upper fibres of the transverse tongue muscle (Hardcastle 1976: 96, 100–106, 134–137); at the low periphery of the spectrum we note that the apical palatal (retroflex) sibilant (as in Toda) has a similar tongue shape to the Spanish anterior sibilant but requires a larger movement of the tongue tip from a schwa-like position towards the palate and a stronger involvement of the upper longitudinal tongue muscle.

In summary, our knowledge of the spectral mean continuum and of the articulatory effort of sibilants is incomplete. This does not preclude the probability that the spectral mean continuum and articulatory effort really exist in human speakers and help shape sibilant inventories, as is strongly suggested by the cross-linguistic observations discussed in this section. We therefore assume both the continuum of the spectral mean and

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

the hypothesis of peripheral articulatory effort. The fact that these hypotheses will make our model work can be regarded as further evidence in favour of them (if we may be allowed a little circularity).

In the following two sections, we show how the dispersion of sibilant inventories can be modelled without dispersion constraints. We illustrate this for English, a language with two sibilants, and for Polish, a language with three sibilants that show chain-shift effects, in §§5 and 6 respectively.

5 The English two-sibilant inventory

English has two sibilants, an alveolar /s/ and a postalveolar /ʃ/. In this section we first show how the English auditory environment leads us to predict the properties of an optimal English OT listener. We describe in detail how an English learner could come to be an optimal listener of her language, and perform a computer simulation showing that a virtual English learner arrives at a pronunciation that matches that of her environment. Having thus shown that in our simulations English is a stable language, we next show that a language with the exaggerated sibilant inventory /s̪/–/ʂ/ is not stable but will instead turn into English within three generations. We do the same for the skewed and confusable inventory /s̪/–/s/.

5.1 The English auditory language environment

For simplicity we assume here that the auditory difference between the two English sibilants is wholly caused by a difference in their spectral means. The spectral mean of /s/, for example, will vary between speakers, vowel environments and acoustic and auditory conditions, as well as between replications by the same speaker. A listener will be able to normalise away some of this variation, but not all of it. Since the non-normalisable variation has multiple sources, the remaining distribution of spectral means for all the /s/ tokens that a listener hears is likely to have a bell-like shape, perhaps similar to the Gaussian shape in Fig. 6. We fix the average spectral mean of all normalised /s/ tokens in the listener's language environment at a slightly arbitrary value of 30·5 Erb (6656 Hz). Likewise, we assume for /ʃ/ a similar Gaussian distribution with an average spectral mean of 25·5 Erb (3478 Hz).[8] For both bell-shaped curves in Fig. 6 we choose a characteristic width (standard deviation) of 1·13 Erb.

[8] Jongman *et al*. (2000) report the values of 6133 and 4229 Hz respectively (for causes described in note 7 above, Gordon *et al*. 2002 typically report differences of only 400 Hz between /s/ and /ʃ/ for the languages they investigate). As noted above, measurement methods for spectral means have not yet been standardised and are therefore difficult to compare across sources. The values used here are therefore based on informal measurements of English and German which employ the same method as Zygis & Hamann (2003) used for Polish (high sample rate; power of 2·0).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

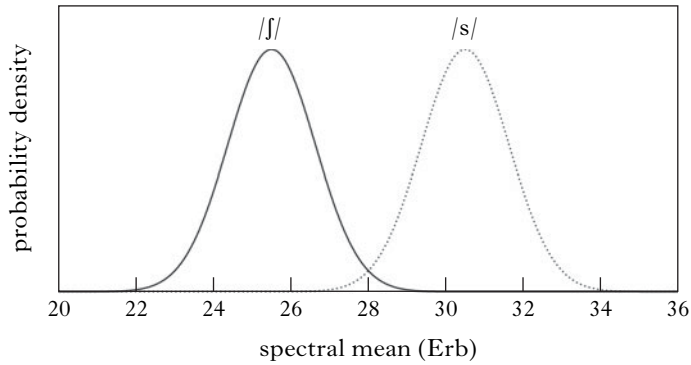


Figure 6
Distributions for tokens of the English sibilants /ʃ/ and /s/ along the spectral mean.

The two distributions in Fig. 6, then, are the material that an adult listener has to work with, and that an infant learner has to make sense of. We assume somewhat arbitrarily that relevant spectral mean values for sibilants universally lie between 20·0 Erb (1789 Hz) and 36·0 Erb (16177 Hz).

5.2 The adult as an optimal listener

We explicitly assume that adult English listeners have learned to minimise the probability of misunderstanding, i.e. that they have become OPTIMAL LISTENERS. To minimise the probability of misunderstanding, the optimal listener must show MAXIMUM-LIKELIHOOD BEHAVIOUR (Fisher 1922), i.e. she must classify an auditory event as the category that is most likely to have been intended by the speaker. For sibilants, the optimal English listener's decision criterion must be the frequency where the two curves in Fig. 6 intersect: she must perceive all tokens with a spectral mean below 28·0 Erb (4765 Hz) as /ʃ/, and all tokens above 28·0 Erb as /s/.

To model this correlation of spectral mean values and surface categories, we follow Escudero & Boersma (2004) in using a complete set of arbitrary cue constraints, which connect 'all' spectral mean values between 20·0 and 36·0 Erb to both sibilant categories. The constraints are summarised in (1).

(1)

*[20·0 Erb]/s/	*[20·0 Erb]/ʃ/
*[20·1 Erb]/s/	*[20·1 Erb]/ʃ/
*[20·2 Erb]/s/	*[20·2 Erb]/ʃ/
…	…
*[35·8 Erb]/s/	*[35·8 Erb]/ʃ/
*[35·9 Erb]/s/	*[35·9 Erb]/ʃ/
*[36·0 Erb]/s/	*[36·0 Erb]/ʃ/

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

For perception, the constraint *[20·0 Erb]/s/ can be read as 'an auditory spectral mean of [20·0 Erb] should not be perceived as the surface phonological category /s/'; for phonetic implementation, the constraint can be read as 'a surface form /s/ should not be realised with a spectral mean of [20·0 Erb]'. For simplification, we discretise the spectral mean range into 161 steps of 0·1 Erb. Although the cochlea, the auditory nerve and the brain perform much more granular discretisations, a granularity of 0·1 Erb is fine enough to oversample any of the three local effects that we will discuss (see §§5.3 and 7.5 for arguments and proof that this discretisation is legitimate and consistent). Combining these 161 spectral mean values with the two sibilant categories results in a total of 322 cue constraints.

In connecting every possible spectral mean frequency between 20·0 Erb and 36·0 Erb to both sibilant categories, the cue constraints in (1) are very different from usual OT constraints, which either tend to express typological trends directly (e.g. markedness constraints) or tend to express universally preferred relations (e.g. faithfulness constraints). That is, the set in (1) has no preference for certain spectral mean values, for certain sibilants, or for connecting certain spectral mean values to certain sibilants, i.e. the constraints are not restricted to actually occurring spectral mean values or to actually occurring combinations of spectral means and sibilant categories. It is the *ranking* of these constraints that will have to be responsible for making the correct connections for English. From Fig. 6 we see, for example, that an optimal listener of English should perceive [26·6 Erb] as /ʃ/. A possible ranking that achieves this perception is given in the perception tableau in (2).

(2) *A perception tableau for classifying tokens with a spectral mean in English*

[26·6 Erb]	*[26·5]/s/	*[26·6]/s/	*[26·7]/s/	*[26·7]/ʃ/	*[26·6]/ʃ/	*[26·5]/ʃ/
a. /s/		*!				
☞ b. /ʃ/					*	

In tableau (2) we have given all possible candidates (only two, because we only consider the English sibilants), but we have restricted ourselves, for reasons of space, to only six of the 322 cue constraints. We can do this because 320 of our constraints have no preference for or against perceiving [26·6 Erb] as /s/ or /ʃ/. Tableau (2) illustrates that the correct classification of [26·6 Erb] relies solely on the ranking *[26·6]/s/ ≫ *[26·6]/ʃ/. In the same way, we can establish rankings for the 160 remaining spectral mean values, e.g. *[31·4]/ʃ/ ≫ *[31·4]/s/; the constraint *[28·0]/s/ will be ranked at approximately the same height as *[28·0]/ʃ/. We have now shown that the 322 constraints can be ranked in such a way that we can model an optimal listener for the distributions of Fig. 6.

However, such a language-specific ranking of 322 constraints is rather uninformative when it comes to predicting what kinds of

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

sibilant-categorisation strategies are possible in the languages of the world. For one thing, the set of 322 constraints seems to generate an incorrect factorial typology: the traditional OT idea of factorial typology would predict the existence of languages with completely undispersed categories, and even the existence of categories with massively non-contiguous auditory correlates.[9] Such languages have not been shown to exist. Apparently, the idea of factorial typology must be incomplete, if cue constraints are to be the correct way to model perception. A general explanation for apparent gaps in the factorial typology is the idea that if UG allows grammars that are not attested in reality, then those grammars might be unlearnable or unstable over the generations (Boersma 2003b). We will therefore model both the acquisition and the evolution of these grammars, and show that undispersed and discontinuous categories are unstable over multiple generations of learners. The following subsection starts by illustrating how the acquisition of English sibilants is modelled with a simple learning procedure and algorithm.

5.3 Learning to become an optimal listener of English

We describe here the situation when a child already has correct lexical representations, but not yet an adult-like prelexical perception. That is, she already knows which lexical items have underlying |ʃ| and which have underlying |s|, but she does not know in the adult-like way of §5.2 what spectral mean values occur with which of the two surface sibilants.

During this acquisition period, the child will receive many tokens of /ʃ/ and /s/ drawn from the distributions in Fig. 6. Sometimes she will make a mistake, as in (3), which occurs when an adult speaker talking to the child pronounces an intended |ʃ| with a reasonable spectral mean of 26·6 Erb.

(3) *A learner's perception tableau with reranking of cue constraints*

[26·6 Erb]	*[26·5]/s/	*[26·7]/ʃ/	*[26·6]/ʃ/	*[26·5]/ʃ/	*[26·6]/s/	*[26·7]/s/
☞ a. /s/					← *	
✓ b. /ʃ/			*! →			

At this arbitrarily chosen point during her acquisition period, the child of tableau (3) happens to have the non-optimal ranking *[26·6]/ʃ/ ≫ *[26·6]/s/ and therefore perceives the incoming auditory event [26·6 Erb] as the sibilant category /s/, as indicated by the pointing finger. However, the speaker had intended to transmit the lexical symbol |ʃ|, and the semantic and pragmatic context may lead the child's comprehension system to

[9] For instance, the ranking {*[24·0]/s/, *[25·0]/ʃ/, *[26·0]/s/, *[27·0]/ʃ/} ≫ {*[24·0]/ʃ/, *[25·0]/s/, *[26·0]/ʃ/, *[27·0]/s/} describes a language in which [24·0 Erb] is perceived as /ʃ/, [25·0 Erb] as /s/, [26·0 Erb] again as /ʃ/ and [27·0 Erb] again as /s/. For a simulation of the learnability of such non-contiguous categories, see §6.4.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

realise this (perhaps because the recognised lexical item was *sheep*). As a result, the child's lexicon can subsequently act as a 'teacher' or 'supervisor' and 'tell' her that she should have perceived /ʃ/ instead of /s/. This new knowledge by the child is indicated by the check mark in the tableau.

When a perception tableau such as (3) contains the child's own winning form (3a) as well as a form that she considers correct (3b), and the two forms are different, the child can conclude that she has made a mistake. As a result, she can take action by taking a 'learning step'. A good strategy for executing a learning step is the GRADUAL LEARNING ALGORITHM (Boersma 1997, Boersma & Hayes 2001), which is indicated by the arrows in the tableau: all constraints favouring the correct category are moved up, and all constraints favouring the child's own 'incorrect' winner are moved down.[10] As a result, the two cue constraints for the value [26·6 Erb] will be re-ranked slightly in the direction of the arrows. This process is called LEXICON-DRIVEN PERCEPTUAL LEARNING (Boersma 1997, Escudero & Boersma 2004).

The arrows in (3) represent small steps along a continuous scale of ranking. After having heard a number of [26·6 Erb] events that should have been perceived as /ʃ/, the learner will have swapped the rankings of *[26·6]/ʃ/ and *[26·6]/s/. From then on, the learner will perceive the auditory form [26·6 Erb] correctly as the category /ʃ/. The same successful learning applies to all other auditory forms for which one of the curves in Fig. 6 is close to zero, i.e. for all forms below approximately [26·7 Erb] or above approximately [29·3 Erb]. It now becomes clear why we called the Gradual Learning Algorithm a 'good strategy': the learner has become a

[10] A reviewer suggests that this algorithm might *not* be a good strategy, referring to Pater's (2008) proof that there are languages (even without variation and without hidden representations) that can be generated by a total ranking in OT but cannot be learned by the Gradual Learning Algorithm (GLA). Our general answer to this is that the ultimate correct learning algorithm for modelling language will be an algorithm that makes the same 'mistakes' as human learners do. For OT, this means that an algorithm is good if it predicts the same gaps in the factorial typology as real languages have (Boersma 2003b). Unfortunately, our specific case is less interesting: there is always exactly one constraint that goes up and one constraint that goes down, and there are no dependencies between different auditory inputs; this is a situation that is trivially handled by formula (15.39) in Boersma (1998: 344). If one is worried that cases with multiple auditory continua do suffer from misconvergences with the GLA for OT, one may try instead to use the GLA for noisy HG (Boersma & Escudero 2008, based on the HG learning algorithm of Soderstrom *et al*. 2006), which does have a convergence proof for cases without variation (Boersma & Pater 2008), or one may use the stochastic gradient ascent algorithm for Maximum Entropy grammars (Jäger, to appear), which has a convergence proof even for cases with variation (Fischer 2005); these algorithms work correctly for our case (§7.6). In order to underline the restricted relevance of algorithm convergence, we would like to point out that no known OT, HG or Maximum Entropy learning algorithm is guaranteed to converge correctly for the arguably more realistic cases with hidden representations (Tesar & Smolensky 2000, Boersma 2003b, Apoussidou 2007, Boersma & Pater 2008). Finally, the whole point of the present paper is that non-optimally dispersed inventories are not completely learnable; the total perception-production learning procedure therefore *must* fail, as it does both in our constraint-based simulations and in Wedel's exemplar-based simulations.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

maximum-likelihood listener for those auditory values. In the region where the curves in Fig. 6 overlap, something slightly different happens: the listener will necessarily continue to make some mistakes in prelexical perception, simply because e.g. an incoming token of [27·5 Erb] was intended as |ʃ| 75% of the time, but as |s| 25% of the time, as the curves show. In Stochastic Optimality Theory (Boersma 1997, Boersma & Hayes 2001), where the ranking of every constraint is subject to a bit of additive noise at evaluation time, the listener is likely to vary her perceptual decisions. The Gradual Learning Algorithm then leads to a situation of PROBABILITY MATCHING: the learner will end up in an equilibrium situation in which she perceives [27·5 Erb] as /ʃ/ 75% of the time, but as /s/ 25% of the time. This works because the constraint *[27·5]/s/ will end up being ranked just above *[27·5]/ʃ/. This strategy of probability matching in perception, which is nearly as good as the maximum-likelihood strategy, has been shown to emerge automatically from the lexicon-driven learning mechanism both for one-dimensional continua (Boersma 1997) and for two-dimensional continua (Escudero & Boersma 2003).

There is one technical correction that has to be made. The above description of the learning algorithm involves changing the ranking of only one cue constraint, as if constraints did not influence their neighbours. A more realistic view of the auditory continuum has to take into account the fact that the properties of the basilar membrane in the inner ear are such that when a hair cell at a certain frequency is excited, several hair cells within a frequency distance of about 1 Erb will also be excited (Moore & Glasberg 1983). This correlation between the firings of adjacent hair cells is reflected in the tonotopic map in the auditory cortex (Romani *et al.* 1982), and presumably in whatever neural device determines the auditory spectral mean. If we acknowledge this mutual correlation between adjacent spectral means and therefore assume a correlation between adjacent cue constraints, the learning algorithm must push bell-shaped dents into the shape of the cue-constraint curves. Following the simulations by Boersma (1997: 51), we implement these dents as Gaussians with a standard deviation (an 'effective resolution') of 0·5 Erb. We have seen, for example, that the ranking of the constraint *[26·6 Erb]/s/ rises by 0·01 in (3). Our vicinity correction now implies that the constraints *[26·5 Erb]/s/ and *[26·7 Erb]/s/ will rise by 0·0098, the constraints *[26·4 Erb]/s/ and *[26·8 Erb]/s/ will rise by 0·0092, and so on; the constraints *[25·6 Erb]/s/ and *[27·6 Erb]/s/ will still rise by $0{\cdot}01 \exp(-0{\cdot}5(1{\cdot}0/0{\cdot}5)^2) = 0{\cdot}0014$. The non-zero effective resolution guarantees a consistent discretisation of the spectral mean continuum by ensuring that the simulation results will be independent of the granularity as long as the spacing (which is 0·1 Erb in this paper) is at least several times smaller than this resolution (see §7.5).

In order to see exactly what happens in perception, and to be able to predict the effect of lexicon-driven perceptual learning on phonetic implementation, we cannot limit ourselves to the present description of the learning mechanism. Instead, we have to *compute* what listeners-speakers

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

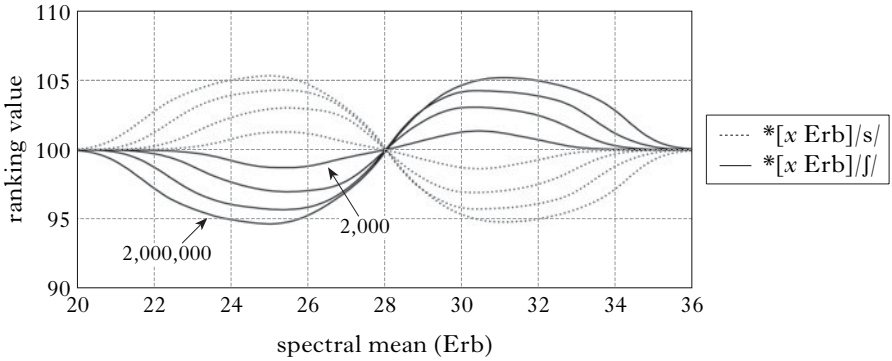


Figure 7

A virtual learner acquiring the perception of the two sibilant categories of English: perception grammars after 2,000, 20,000, 200,000 and 2,000,000 input tokens.

will do, and since any case involving more than two constraint curves is impossible to solve analytically, a computer simulation is the designated route.

5.4 Simulation of the acquisition of English perception

The learning of an English perception grammar can be simulated by computer. Our virtual learner has the 322 cue constraints in (1), and starts out having them all ranked at the same height of 100·0; as a result, the learner has an initial state where she perceives every incoming spectral mean value as /s/ 50 % of the time and as /ʃ/ 50 % of the time. At the same time we assume that this initial virtual learner already has correct lexical representations for all words with /s/ and /ʃ/. Although this combination of fully random prelexical perception and perfect lexical storage is obviously unrealistic, a more realistic modelling is very likely to exhibit effects that work in the same directions as the ones we will find here, although their size may differ (see §7.3 for discussion).

During the simulated acquisition period, our learner hears 1 million /ʃ/ and 1 million /s/ tokens in random order and with spectral mean values that are randomly drawn from the distributions in Fig. 6. Each token is also labelled as /s/ or /ʃ/ by the learner's lexicon. When hearing a token, the learner will perceive it into either category. In our simulations we set the standard deviation of Stochastic OT's EVALUATION NOISE (per tableau random variation in ranking) to a constant value of 2·0.

It can happen that the learner's perceived category is identical to the category the lexicon tells her she should have perceived. In such a case, the learner does not change her grammar. But if the perceived category is different from the one her lexicon says is correct, our learner changes the ranking of her cue constraints according to the scheme in tableau (3). The

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

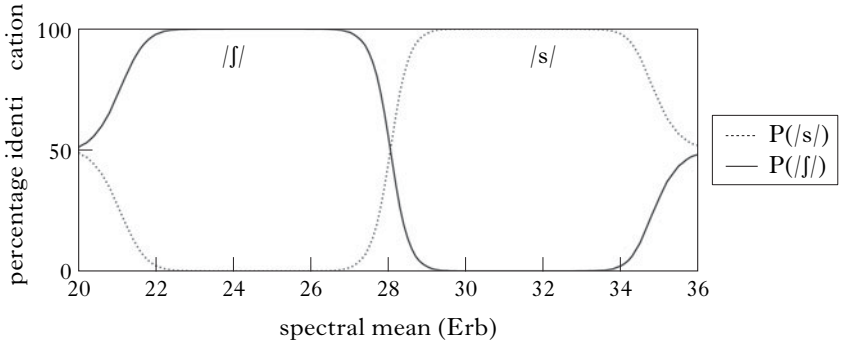


Figure 8
Identification curves for the virtual listener of Fig. 7 (final state).

Gradual Learning Algorithm is here taken to have a constant PLASTICITY (re-ranking step) of 0·01.

While listening to the 2 million tokens, our learner's cue constraints gradually move away from 100·0. The final perception grammar and three intermediate stages are shown in Fig. 7, where the constraints for the sibilant /ʃ/ are connected by a solid curve, and those for /s/ by a dotted curve. We can see, for instance, that for a spectral mean value of 24·0 Erb (2896 Hz), the dotted curve ends up having a ranking value of 105·0, and the solid curve one of 95·0. This means that the cue constraint *[24·0 Erb]/s/ ends up being ranked much higher than *[24·0 Erb]/ʃ/, and thus the token [24·0 Erb] is most often perceived as /ʃ/ in the final perception grammar (not always, because of the evaluation noise). In general, the lowest curve determines which category is perceived most often at any specific spectral mean value in Fig. 7.

A complete OT analysis of a problem generally involves two parts, namely a description of the constraint ranking and a description of what outputs the grammar assigns to all of its inputs. While the ranking of all 322 constraints is given in Fig. 7, a description of the workings of the grammar would involve giving 161 input–output pairs, perhaps in the form of 161 perception tableaux that are analogous to tableau (2). And even these 161 tableaux would not suffice, because as a result of the evaluation noise of Stochastic OT it is fully possible that the same spectral mean value is sometimes perceived as /ʃ/, and sometimes as /s/. A full account of how the 161 spectral mean values are handled by the perception grammar therefore involves giving for each of the 161 spectral mean values the probability that it is perceived as /ʃ/ and the probability that it is perceived as /s/. An estimate of these is given in Fig. 8. This figure has been computed by running each of the 161 spectral mean values through the final perception grammar 100,000 times (with an evaluation noise of 2·0, as during learning), and noting how often the output was /ʃ/; dividing each of the 161 results by 100,000 yields an estimate of the probability that

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

each spectral mean is classified as /ʃ/.[11] The probability of classifying it as /s/ is estimated in an analogous way.

In Fig. 8 we see that the probability of perceiving /ʃ/ is greatest for those spectral values *x* for which the curve of *[*x* Erb]/ʃ/ lies furthest below the curve *[*x* Erb]/s/ in Fig. 7. This is a general property of Stochastic OT: the further constraint A is ranked above constraint B, the smaller the chance that B will overcome A at evaluation time.

Figure 8 confirms that the Gradual Learning Algorithm leads to optimal perception if the learner is given sufficient information. In the regions where she has heard a large number of spectral mean tokens (i.e. between approximately 23 and 33 Erb, as can be seen from inspecting Fig. 6), Fig. 8 shows that the learner has become an optimal (probability-matching) listener. For instance, the spectral mean of 27·5 Erb has (according to Fig. 8) a 75% probability of being perceived as /ʃ/ and a 25% chance of being perceived as /s/; hence, the odds of perceiving 27·5 Erb as /ʃ/ are three times as high as the odds of perceiving it as /s/; this ratio of 3 to 1 corresponds exactly to the relative heights of the two curves in Fig. 6 at [27·5 Erb].

However, the learner does not become optimal if she is given too little information. In the left periphery of the auditory space (around 21 Erb), the probability that such a spectral mean value was intended as /ʃ/ is very much higher than that it was intended as /s/, and this would predict that a probability-matching listener perceives such spectral mean values as /ʃ/ 100% of the time. Nevertheless, Fig. 8 shows that in this region the learner's perception grammar varies between perceiving /ʃ/ and /s/. This imperfection arises because the learner has not heard enough peripheral tokens to drag the two curves apart in this region.[12] The same holds for the right periphery of the auditory space (around 35 Erb).

It is interesting to see for which spectral mean values the probabilities are maximal. The probability-matching criterion would predict that the probabilities would be maximal in the regions near 20 and 36 Erb, but the scarcity of such tokens has moved the point of maximal separation quite far toward the centres of the two categories, i.e. towards 25·3 and 30·7 Erb. The spectral mean values for which the two curves are furthest apart are approximately 24·9 and 31·1 Erb. These locations can thus be said to represent the least confusable spectral mean tokens, and will turn

[11] In this specific case, with only two categories, the probability could be computed directly (rather than just estimated) by the formula in Boersma (1998: 331). However, this would not work for the more complicated case of Fig. 17, where each spectral mean value involves three constraints.

[12] One might think that considerations of auditory distance (21 Erb is closer to the centre of the /ʃ/ category than to the centre of the /s/ category) predict that real listeners always classify [21 Erb] as /ʃ/. However, the uncertainty that our virtual listener displays around 21 Erb may well be realistic: an inspection of Escudero & Boersma (2004: Fig. 3) shows that Southern British English listeners have trouble classifying [ɛ] as an instance of the auditorily closer category /ɪ/ rather than the remoter category /i/.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

out to be relevant when we consider how the listener of Fig. 7 will pronounce /ʃ/ and /s/ herself.

5.5 The near-optimal English listener's preferred production: the prototype effect

In a bidirectional model of phonetics, the cue constraints and rankings used by the listener in perception are also used in her phonetic implementation. When implementing an articulation, the cue constraint *[20·0 Erb]/s/ is read as 'an /s/ should not be produced with a spectral mean of [20·0 Erb]', and so on.

In (4) we see how the virtual learner of § 5.4 would now produce an /s/, if only the cue constraints and rankings of Fig. 7 were involved.

(4) *A production tableau with cue constraints only*

/s/	*30·6 /s/	*30·7 /s/	*30·8 /s/	*31·5 /s/	*30·9 /s/	*31·4 /s/	*31·3 /s/	*31·0 /s/	*31·2 /s/	*31·1 /s/
a. [30·6 Erb]	*!									
b. [30·7 Erb]		*!								
c. [30·8 Erb]			*!							
d. [30·9 Erb]					*!					
e. [31·0 Erb]								*!		
☞ f. [31·1 Erb]										*
g. [31·2 Erb]									*!	
h. [31·3 Erb]							*!			
i. [31·4 Erb]						*!				
j. [31·5 Erb]				*!						

The candidate [31·1 Erb] in (4) wins because the curve of the cue constraints for /s/ in Fig. 7 is lowest at this value (which is also where the two curves are maximally separated). If we compare this phonetic output to the token with the highest frequency in the distribution of Fig. 6, which has 30·7 Erb, we can see that our speaker produces an /s/ that has shifted slightly by 0·4 Erb from the /s/ that is most commonly produced by her surroundings.

We now show that the shift is actually larger than the 0·4 Erb seen in (4). As a result of the evaluation noise of Stochastic OT, the winner in (4) will not always be 31·1 Erb. Instead, the winner will be the spectral mean value whose cue constraint (for /s/) happens to be lowest-ranked when evaluation noise is added to the ranking of each constraint. If we apply the same evaluation noise as in perception, namely with a standard deviation of 2·0, all spectral mean values with a ranking not much higher than that of *[31·1 Erb]/s/ are also likely to win. In Fig. 7 we see that the curve of *[*x* Erb]/s/ has a low plateau in the whole region between, say, 30·0 and 32·8 Erb, and these are all quite likely to win. The right-hand side of Fig. 9

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

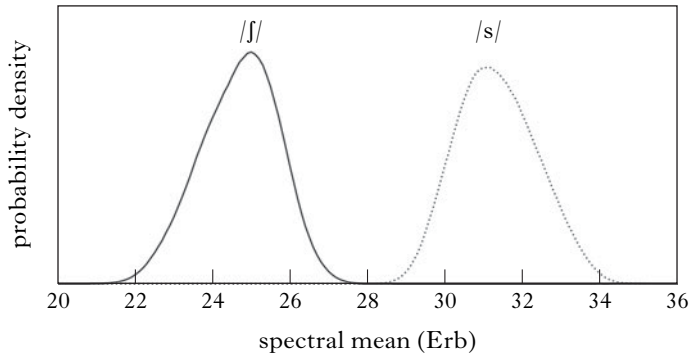


Figure 9
Output distributions of the inverted perception grammar (distributions of 'prototypes').

shows the distribution of spectral values that we obtain by running the category /s/ 10 million times through production tableaux with the rankings of Fig. 7. The average spectral mean of the /ʃ/ category is 24·69 Erb, that of the /s/ category 31·35 Erb. The size of the shift (from the original 25·5 and 30·5 Erb) towards the periphery is therefore approximately 0·83 Erb.

What is behind this shift? From the discussion at the end of § 5.4 we can infer that a probability-matching learning algorithm works in such a way that the constraints for values that are least likely to have been intended by the speaker as anything other than the dominant category in that region tend to be lowered furthest. Very roughly, the learning algorithm causes cue constraints to end up ranked lowest in auditory regions where the learner has heard the largest number of least confusable tokens. Teleologically speaking, our speaker prefers to produce an /s/ that is more peripheral than the average token that she has heard herself, because the auditory distance of such a token from the competing /ʃ/ is larger. What we observe here is the PROTOTYPE EFFECT in OT (Boersma 2006). This phenomenon has been attested in experiments in the laboratory: when speakers of a language are asked to choose the best auditory instance of a sound category, they choose a more peripheral token than they would actually produce themselves (Johnson *et al.* 1993), apparently because they choose the token that is least likely to present any other category than the one requested. It is important to note that although this explanation is couched in teleological terms (as if the listener-speaker knows about this), the underlying mechanism is not explicitly goal-oriented at all (and does not even know about auditory distance): the prototype effect occurs in OT simply because people employ in production the same cue constraint rankings that have optimised their prelexical perception.

The conclusion must be that learners will end up preferring more peripheral tokens than they have heard on average in their language environment. This would predict a sound shift if real learners really did

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

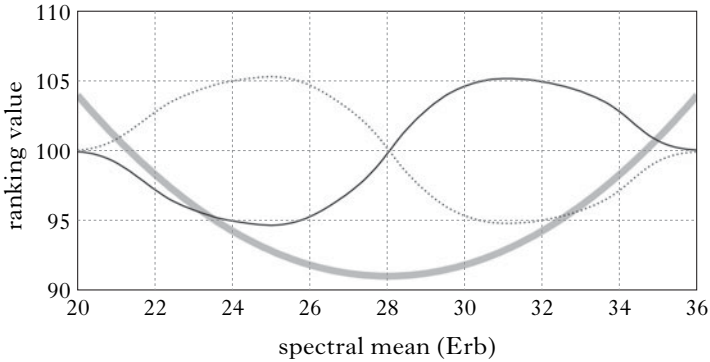


Figure 10
A production grammar for the two sibilant categories in English.

that. But the phonetic implementation process of real speakers is not determined solely by their cue constraints. Articulatory considerations will be seen to keep the prototype effect within bounds.

5.6 Really speaking involves more constraints: the articulatory effect

Something is wrong with the assumption in §5.5 that phonetic implementation involves cue constraints alone. In real production the speaker is restricted by articulatory constraints as well. For instance, we can define the constraint *[31·2 Erb] as the articulatory constraint whose ranking reflects the articulatory effort associated with producing a spectral mean of 31·2 Erb. In the following we assume the independently plausible hypothesis (§§2.1, 4) that articulatory effort is minimal (though definitely not zero) for central auditory values of the spectral mean, and that more peripheral auditory values are harder to produce. If this hypothesis is correct (it at least explains the central tendency in single-sibilant inventories in §2.1), the articulatory constraints for peripheral auditory values must be ranked higher than those for more central auditory values.

In the production of a sound, articulatory constraints interact with the cue constraints. This is illustrated with the production grammar in Fig. 10, where articulatory constraints are represented with a thick grey curve, following a parabolic shape reminiscent of effort curves found in biophysics (Hoyt & Taylor 1981). The cue constraints have the same rankings as in the perception grammar in Fig. 7, and are again represented as solid and dotted curves. Following Kirchner (1998), we assume for simplification that the ranking of articulatory constraints is fixed and not influenced by language-specific learning.[13]

[13] A more realistic model would involve articulatory learning, which should lower the ranking of articulatory constraints for spectral mean values that the speaker has been practising (Boersma 1998: ch. 14).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

The interactions of articulatory and cue constraints in production become clear in the tableau in (5), where our learner tries to articulate /s/.

(5) *A production tableau with cue constraints and articulatory constraints*

/s/	*31 ·2	*31 ·1	*31 ·0	*30 ·9	*30 ·6 /s/	*30 ·8	*30 ·7 /s/	*30 ·7	*30 ·8 /s/	*30 ·6	*30 ·9 /s/	*31 ·0 /s/	*31 ·2 /s/	*31 ·1 /s/
a. [30·6 Erb]					*!					*				
☞ b. [30·7 Erb]							*	*						
c. [30·8 Erb]						*!			*					
d. [30·9 Erb]				*!							*			
e. [31·0 Erb]			*!									*		
f. [31·1 Erb]		*!												*
g. [31·2 Erb]	*!												*	

The candidate [31·1 Erb], which was the winner in the tableau without articulatory constraints in (4), no longer wins, because its articulation involves more effort than that of [30·7 Erb], which is the new winner. Loosely speaking, candidates below [30·7 Erb] are too indistinctive and candidates above [30·7 Erb] are too hard (note that in this tableau *[30·8 Erb] outranks *[30·7 Erb]/s/ despite the fact that in the region of 30·7 and 30·8 Erb the articulatory curve in Fig. 8 lies below the cue curve for /s/; there is no contradiction: in stochastic OT, constraints that are as closely ranked as these two will be ranked in the opposite order in a non-negligible fraction of the evaluations).

While the bidirectional use of cue constraints causes the categories to drift apart auditorily (§5.5), the presence of the articulatory constraints checks this expansion and drives the production distributions back towards the centre of the spectral mean continuum. Figure 11 shows the production distributions, estimated by running each sibilant category 10 million times through the grammar of Fig. 10 with an evaluation noise of 2·0.

The average spectral mean of the /ʃ/ category is 25·31 Erb, and that of the /s/ category is 30·70 Erb. When we compare this with the averages of Fig. 6 (25·5 and 30·5 Erb) and Fig. 9 (24·69 and 31·35 Erb), we conclude that the articulatory constraints have effectively prevented the categories from drifting apart auditorily: our English learner produces nearly the same average spectral means as her parents. Apparently, the articulatory effect has cancelled out the prototype effect. This balance of powers was first noted and modelled by Boersma (2006).

There is an important difference between the distributions in Fig. 6 and Fig. 11. In Fig. 6 the standard deviations of the two distributions are 1·13 Erb, and in Fig. 9 they are 1·03 and 1·06 Erb. In Fig. 11, however, the standard deviations are only 0·85 and 0·86 Erb: the distributions are narrower. Apparently, the articulatory effect has caused, in addition to the

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

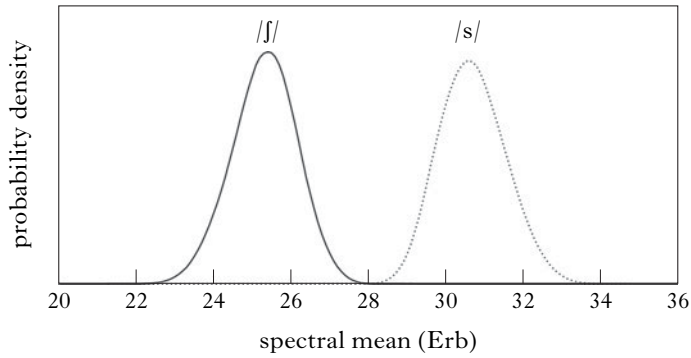


Figure 11
The production distributions for the two sibilant categories in English.

shift, an ENTRENCHMENT of the produced distribution when compared with the distribution heard. This sharpening is an important property of our model, which will be seen to help to make the distributions stable over the generations (§5.7) and to allow the model to be compared favourably to other models (§7.4).

The near-exact cancelling of the articulatory and prototype effects is a direct result of our choice of parameters: knowing that English has a stable sibilant system, we chose the height of the articulatory curve to be the one in Fig. 10, not a higher or lower one. In this sense, the parameters of the model have been established on the basis of language data. Therefore we have to question the equal sizes of the prototype effect and the articulatory effect, and will do so in §§5.8 and 5.9. First, however, we show that the language simulated in §§5.4–5.6 is indeed stable over the generations.

5.7 Simulating sound change: a stable language

Are the average spectral means of the English sibilants, namely 25·5 and 30·5 Erb, indeed stable over the generations, or do they slowly drift? We test this by simulating the acquisition of the two sibilants over nine more generations.

Some care has to be taken in how the produced spectral mean values of the first generations of learners are fed to the second generation. We cannot transmit the distribution of Fig. 11 directly from speaker to learner, because the only variation in the spectral mean values of Fig. 11 is due to decision noise in production. That is, the standard deviation of 0·85 Erb only reflects the evaluation noise in the production tableaux. The listener will be confronted with this source of variation (and not be able to normalise it away), but also with some non-normalisable between-speaker variation (because real learners will hear multiple speakers), some random variation within the speaker's muscle system (which is independent from the speaker's evaluation noise), the background noise in the air and the

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

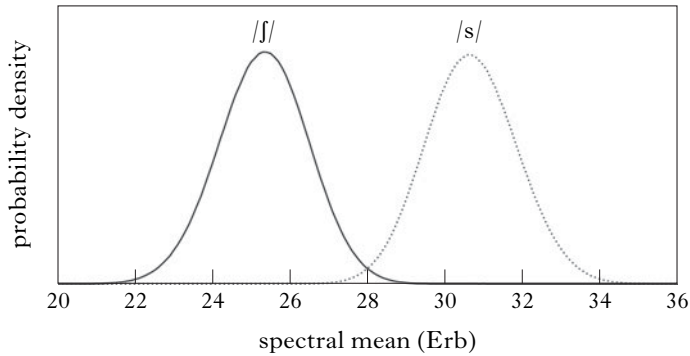


Figure 12
The sibilant environment of the second generation of learners.

noise in the learner's ear. We represent these influences on the spectral mean together as TRANSMISSION NOISE, with a standard deviation of 0·80 Erb. More precisely, the second generation of learners will be presented with spectral mean tokens computed from the grammar in Fig. 10, where to each token produced we add a value drawn randomly from a Gaussian curve with a mean of zero and a standard deviation of 0·80 Erb. The resulting distribution of spectral mean input values for generation 2 is given in Fig. 12, which was estimated by running each sibilant category 10 million times through the grammar of Fig. 10 with an evaluation noise of 2·0 and a transmission noise of 0·80 Erb. The standard deviations are approximately 1·17 Erb.

Our second generation is thus represented by a learner who acquires a perception grammar in the way described in §5.4, on the basis of outputs of generation 1 modified by transmission noise (we again assume that the learner can distinguish the categories from each other). This learner then turns into a speaker by including articulatory constraints, as described in §5.6. We then feed tokens produced by this speaker, modified by 0·80 Erb transmission noise, to a new learner, our third generation. This learner acquires again a perception grammar and a production grammar; her transmitted output is used as input to the fourth learner, and so on.

In total we simulate this acquisition process for 10 learners in a row, with the simplifying assumption that every learner stands for a whole generation of speaker-listeners. Every learner receives 1 million /ʃ/ and 1 million /s/ tokens, and has a plasticity of 0·01 and an evaluation noise of 2·0.

Over the generations, the English spectral mean values stabilise at about 25·40 and 30·60 Erb, and the standard deviations stabilise at about 1·18 Erb. This is shown in Fig. 13, where the black curves connect the single-speaker averages and the grey areas represent the standard deviations of distributions like those in Figs 6 and 12 (i.e. including transmission noise).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

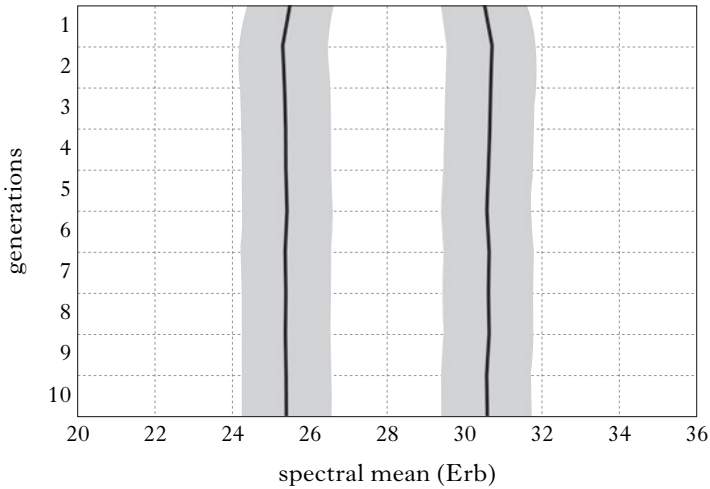


Figure 13
The spectral means of the two English sibilants and their stable learning over 10 generations.

Our simulated English (Fig. 6) turns out to be boringly constant (Fig. 13). There is no real change between the output of generation 1 and the output of generation 10 (except that the distributions are no longer Gaussian, but slightly skewed, as in Fig. 12). The next question is whether a language with different auditory distributions than Fig. 6 is equally stable, or whether it instead drifts away from its initial distributions.

5.8 Simulating sound change: a language with an exaggerated contrast is unstable

Here we simulate the learning of a language with a much more extreme two-sibilant contrast than English, i.e. [s̻] *vs.* [s̺], with spectral means of 23·0 and 33·0 Erb. Except for these initial distributions, everything else (including the shape and height of the articulatory constraint curve of Fig. 10) is the same in this simulation as in the one of §§5.4–5.7.

Figure 14 shows the result. For this 'exaggerated English', the second generation more or less learns the oversized range, although they reduce it to a much more moderate contrast of 24·83 ~ 31·18 Erb. Apparently, the articulatory effect outweighs the prototype effect for this generation; this comes as no surprise, since the articulatory curve in Fig. 10 (in the regions of 24·83 and 31·18 Erb) approaches the cue-constraint curve. The third generation has already shifted the system towards an unmarked articulatorily-perceptually balanced [ʃ] and [s]. Within two generations, the learners have changed the exaggerated English into plain English.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

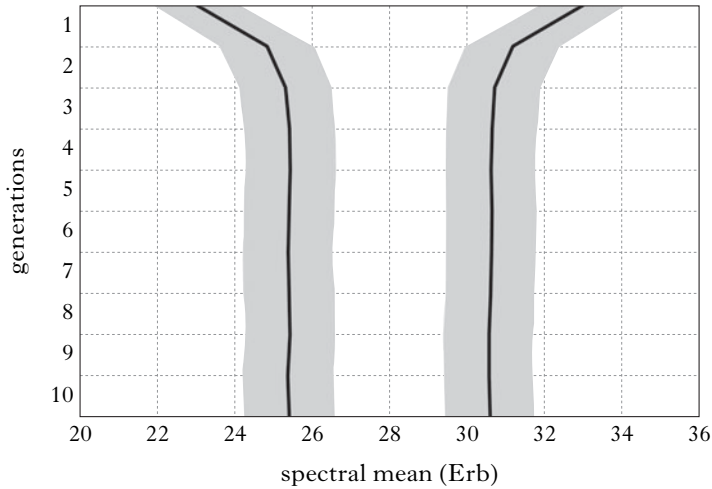


Figure 14
The development of the two sibilants in 'exaggerated English' over 10 generations.

The conclusion is that the articulatory and prototype effects can act independently, and work together to establish, non-teleologically, an optimal balance between articulatory ease and auditory contrast.

5.9 Simulating sound change: a language with a confusable contrast is unstable

If the optimal balance achieved in Figs 13 and 14 is characteristic of two-sibilant inventories in general, then the 25·4 ~ 30·6 Erb inventory should emerge for every possible initial inventory. The two cases that yet have to be investigated in this respect are the case of an initial 'confusable' English, where the categories are closely spaced, and the case of an initial 'skewed' English, where the categories are not positioned symmetrically around 28·0 Erb (as they are in Figs 13 and 14).

The two cases are combined in the simulation of Fig. 15, a language with initial sibilants at 28·0 and 32·0 Erb, which is both skewed (one sibilant has a central, the other a high spectral mean) and relatively confusable (the difference is only 4·0 Erb, rather than the 5·0 Erb of the earlier simulations).

The first generation of learners moves the spectral mean values apart (to 26·8 and 32·2 Erb), immediately reaching an English-like distance of about 5 Erb. Apparently, the prototype effect plays a very fast role here; this confirms the repulsive dynamic nature of the present model and shows that the model will be able to handle phonetic enhancement and chain shifts. The articulatory effect is also seen to play a role, because

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

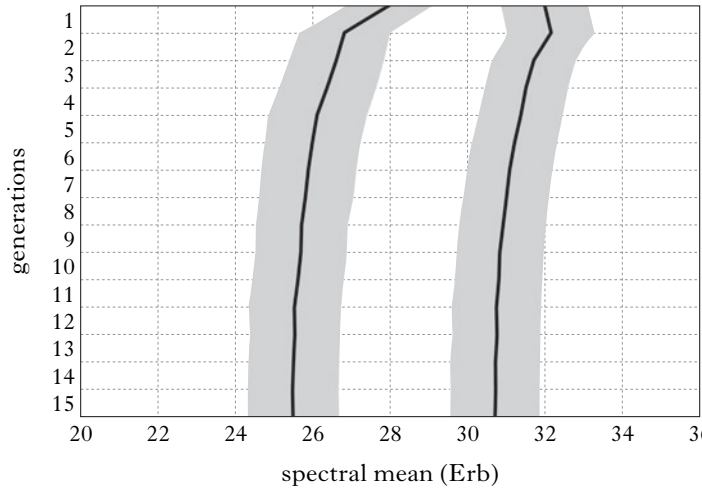


Figure 15
The development of the two sibilants in 'skewed and confusing English' over 15 generations.

the amount the lower category moves down (i.e. to the left in the figure) is larger than the amount the top category moves up. This effect works quite slowly, however: it takes quite a number of generations to obtain a perfectly symmetric inventory. The cause of the slowness is that the upper category 'wants' to move down but that this is hampered both by the repulsive force from the lower category and from the fact that the lower category itself experiences an upward bias from the articulatory effect. Nevertheless, a symmetric English-like inventory is eventually reached.

We can conclude from Figs 13–15 that our model predicts that, independently of the situation in generation 1, the inventory always evolves towards the same two auditory values. In other words, all stable languages with two sibilants are like English (or French, or German).

So we see that optimal dispersion indeed happens. Figs 13–15 all show the effect of the excluded centre (§2.1): the region around 28 Erb is avoided in languages with two sibilants. The next step is to look at other kinds of sibilant inventories: which of the dispersion effects will we see?

6 Larger, smaller and different inventories

In order to find all six dispersion effects discussed in §2.1, the present section discusses inventories with three and four categories (cf. Fig. 5). As special cases we also discuss an inventory with a single category and an inventory containing a non-contiguous category.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

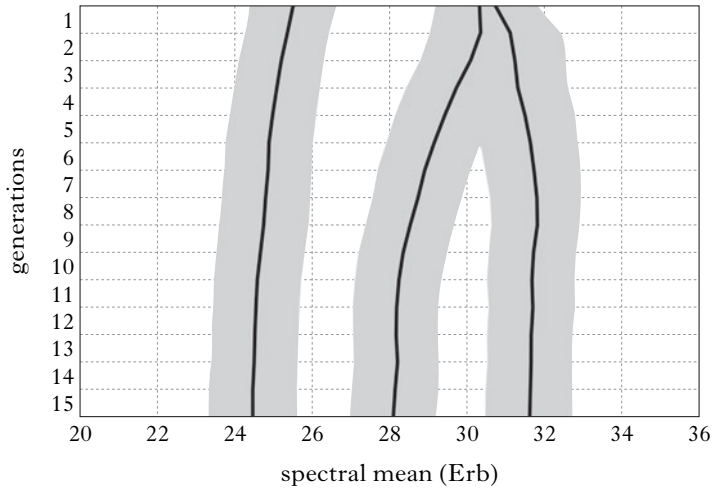


Figure 16
The development of the three sibilants in Polish over 15 generations.

6.1 The Polish three-sibilant inventory: a chain shift

Polish used to have the sibilants /ʃ sʲ s/ (Carlton 1991). If the dispersion principle is correct, and if the spectral mean is the only auditory cue that distinguishes these three sibilants, such an inventory cannot be stable. The simulation in Fig. 16 shows what happens if we start out with a language whose sibilants have spectral means of 25·5, 30·3 and 30·7 Erb (in order to make our point, we have taken values for /sʲ/ and /s/ that are closer together than they probably actually were in medieval Polish).

The first striking phenomenon that Fig. 16 shows is that the /sʲ/ ~ /s/ contrast is phonetically enhanced: the palatalised alveolar (the middle category) lowers its spectral mean beyond 29 Erb, i.e. into the [ɕ] region. This is reported to have happened in real Polish in the 13th century (Stieber 1952, Carlton 1991). The second thing that happens is that the laminal postalveolar (the sound on the left in Fig. 16) shifts down towards the apical postalveolar ('retroflex') [ʂ], which is reported to have happened in real Polish in the 16th century (Rospond 1971, though this sound is usually transcribed as postalveolar /ʃ/; see the discussion in Hamann 2004). Our simulation (when compared with the English simulations) confirms Jones' (2001) proposal that this second shift was caused by the first, i.e. that we are observing a contrast-enhancing push chain here: the /ʃ/ category is shifted down as a result of the approach by the lowering /ɕ/. Furthermore, we can observe the equally contrast-enhancing shift (again proposed by Jones 2001) of the alveolar /s/ towards a more peripheral [ʂ], its present-day location (Puppel *et al*. 1977). This simulation thus explains the present Polish sibilant system /ʂ ɕ ʂ/, which has spectral means very close to the 24·46, 28·08 and 31·59 Erb found here (the three female

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

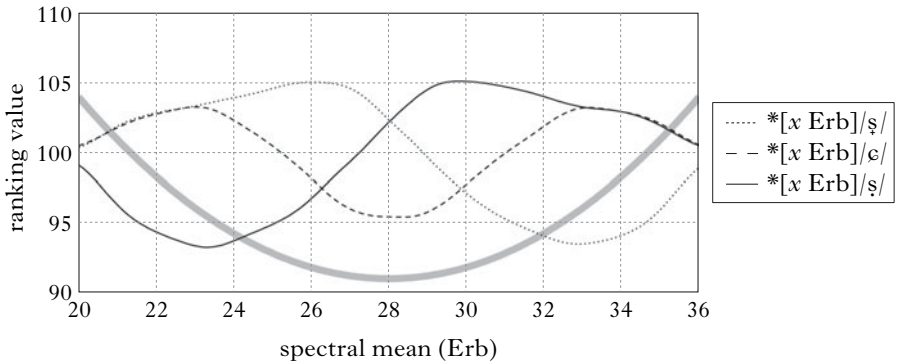


Figure 17
A production grammar for the three sibilant categories in Polish.

speakers recorded for the study by Zygis & Hamann 2003 have average values of 24·40, 27·82 and 31·29 Erb). The grammar acquired by the fifteenth generation in Fig. 16 is given in Fig. 17; the articulatory constraint curve is the same as in Fig. 10. As before, this grammar is bidirectional. When given a certain spectral mean value, the grammar will most likely perceive it as the category for which the cue constraint at that spectral mean value is the lowest of the three. When given a category, the grammar will produce a spectral mean value where both the relevant cue constraint curve and the articulatory constraint curve are low; the average realised spectral mean values are 24·46, 28·08 and 31·59 Erb, as mentioned above. We have thus established an account of the Polish sibilant inventory without using dispersion constraints: only the cue and articulatory constraints of Fig. 17 are necessary (cf. Padgett & Zygis 2003, who did use dispersion constraints for this example).

When we compare Fig. 16 to Fig. 15, we see all the dispersion effects discussed in §2.1. First, the centre category (around 28·0 Erb) is back in the picture. Second, the dispersion in Fig. 16 is even: both pairs of adjacent categories are removed from each other by approximately 3·6 Erb. Third, we see that the size of the auditory space is greater when there are three categories (7·1 Erb) than when there are two (5·2 Erb). Fourth, from the widths of the grey areas we see that the stable variation within a category is somewhat smaller when there are three categories (namely 1·12 Erb for the middle category and 1·15 Erb for the peripheral categories) than when there are two (1·18 Erb, according to §5.7). Fifth, we have observed a push chain when a lowering mid category caused the bottom category to lower as well. Sixth, we will observe a drag chain if we remove the bottom category from the final state in Fig. 16: from the simulation in §5.9 we know that the original mid category will move down to 25·4 Erb, allowing the top category to move down from 31·6 to 30·6 Erb. We conclude that we have faithfully modelled all aspects of

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

auditory dispersion by assuming bidirectionality of cue constraints and a U-shaped articulatory effort curve, as in the production grammars in Figs 10 and 17.

6.2 A four-sibilant inventory

The largest inventory we consider is one with four sibilants, like Toda in Fig. 5. Independently of where the centres of the categories lie in the environment of the learners of the first generation, the centres of the categories evolve towards values around 24·4, 26·5, 29·5 and 31·6 Erb, i.e. the distances between the categories are again smaller than in the three-sibilant case, while the total space taken up has again increased (although very little). The spacing between the categories is somewhat greater in the middle than at the edges, probably because at the edges the limiting effect of the articulatory constraints is greater (this can be seen as a more precise formulation of §2.1.3). The final standard deviations are 1·13 Erb for the two outer categories and 1·22 Erb for the two inner categories, i.e. a bit larger than in the three-sibilant case.

We conclude that all dispersion effects identified in §2.1 remain valid in larger inventories (except the unexpectedly high within-category variation in the four-sibilant inventory, for which we have no explanation).

6.3 A one-sibilant inventory

The smallest inventory we consider is one with a single sibilant. The first generation of learners already turns up with a category centred at the centre of the auditory continuum, i.e. 28·0 Erb, and the standard deviation will be and stay a gigantic 3·03 Erb. The cause of this situation is that the learners do not learn: if the only category along the continuum is the unspecified sibilant /S/, they will perceive any spectral mean value as /S/, and therefore never make a mistake. As a result, all cue constraints stay ranked at 100·0, and in production the decision about which auditory value to pronounce is determined partly by which cue constraint happens to be lowest ranked, partly by the articulatory constraints.

The reader may object that the predicted one-shot shift to the centre is unrealistic, yet this is what we predict will happen in the absence of other phonological entities. In practice it will be very difficult to find such a situation: the single (retracted apico-alveolar) sibilant of Iberian Spanish, for instance, must cope with the existence of a rather strident /θ/ in the same language.

6.4 Can non-contiguous categories be learned?

The present model does not involve any representation of auditory distance. That is, the learner represents nothing more than 161 values along the spectral mean continuum, and does not necessarily know that they are ordered in a natural way. For instance, the virtual learner never needs to

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

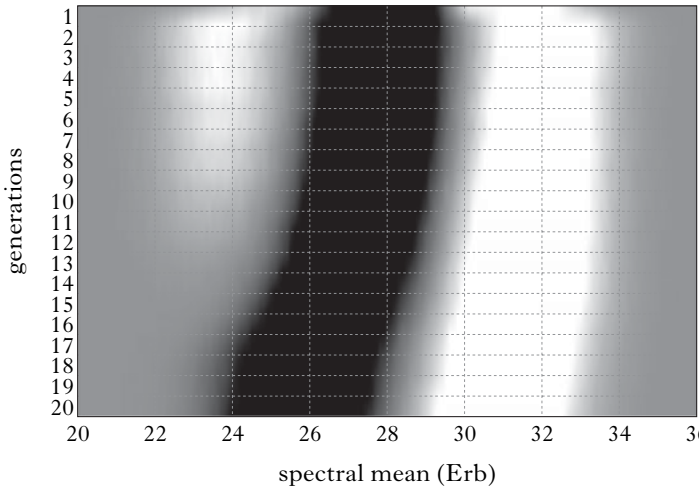


Figure 18

The demise of a bimodally distributed category. Black = /ʃ/, white = /s/.

know that the spectral mean associated with the 37th value along the continuum (i.e. 23·6 Erb) lies in between those of the 27th value (22·6 Erb) and the 47th value (24·6 Erb). As a result, the constraint set of our virtual learner can easily represent an inventory where category 1 has spectral means around 26·5, 28·5 and 30·5 Erb, whereas category 2 has spectral means around 25·5, 27·5 and 29·5 Erb (as was mentioned in note 9). The question naturally arises, however, whether such discontiguous categories are not only representable, but learnable as well.

The answer turns out to be that discontiguous categories are 'semi-learnable': if the learner's environment contains a discontiguous category, then the learner's final category may still be discontiguous, but less so than the one in her environment, and within a number of generations the language will have changed into one with only contiguous categories.

An example is shown in Fig. 18. The initial inventory has a simple (monomodal) category (which we arbitrarily label /ʃ/) centred at 28·0 Erb with the usual relative frequency of occurrence of 50%, plus a bimodal category (which we label /s/) with a peak centred at 31·0 Erb with a relative frequency of 37·5% and a peak centred at 25·0 Erb with a relative frequency of 12·5%; all three peaks have a standard deviation of 1·13 Erb. Thus, both categories are equally likely, and the upper part of /s/ is three times more likely than the lower part.

Figure 18 shows that the following generations continue to associate the /s/ category with two peaks. However, the lower peak shrinks slowly, and has almost disappeared after generation 11. After 20 generations, the English inventory has arisen again.

We conclude that in our model non-contiguous categories are learnable, but not stable over the generations. The ultimate cause of the

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

development into monomodal distributions can be explained with Fig. 18 in the following way. The fact that the upper peak of /s/ is taller than the lower peak causes an asymmetry in the regions of confusion. That is, the region of confusion between the lower peak and /ʃ/ lies somewhat further from the centre of the continuum (namely around 26·0 Erb, which is 2·0 Erb away) than the region of confusion between the lower peak and /ʃ/ does (around 29·5 Erb, which is only 1·5 Erb away). As a result, /ʃ/ will shift down (i.e. to the left), which then pushes the lower peak of /s/ down into an articulatory more effortful region (push chain) and allows the upper peak of /s/ to come down into a less effortful region (drag chain). As a result, the articulatory bias against the lower peak of /s/ becomes greater than the articulatory bias against the upper peak, and this causes a slight preference in production for selecting tokens from the upper peak. This process is self-reinforcing, because it moves the boundaries between the three categories down.

We do not yet know how to assess our predicted relative learnability of bimodal distributions. If there exist mechanisms that cause such distributions to arise, we must be able to observe them in real languages, because such distributions are predicted to take ten generations to become monomodal. Whether actual languages do have this kind of transitory allophony in sibilants remains to be seen.[14]

7 Discussion

Our simulations within a bidirectional model of phonetics realistically predict that a language with one, two, three or four sibilants automatically evolves towards a stable DISPERSED system, i.e. one that has single-peaked categories equally spaced along the auditory spectral mean continuum. The end result of such an evolution is independent of the spectral means of the categories in the first generation: given the number of categories, the resulting final categories will always be monomodal and have the same averages and standard deviations. Another thing our simulations have modelled realistically is the diachronic development of the sibilant inventory of Polish.

Our approach reconciles the standpoint of innocent misapprehension with that of auditory dispersion: speakers are not goal-oriented, but at the same time sound change tends to minimise perceptual confusion. Sound change at the level of the language learner is thus non-teleological, whereas at the abstract level of the observed language it is teleological. In

[14] We might speculate that Dutch is becoming such a case. The introduction of the relatively new and somewhat marginal alveolopalatal sibilant /ɕ/ (usually transcribed as /ʃ/) into an existing inventory consisting solely of an (auditorily equally central) flat laminal alveolar sibilant (usually transcribed as /s/) may cause a split in the population between varieties of /s/ with higher and those with lower spectral means than /ɕ/. Unfortunately, there is no room here to test this speculation.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

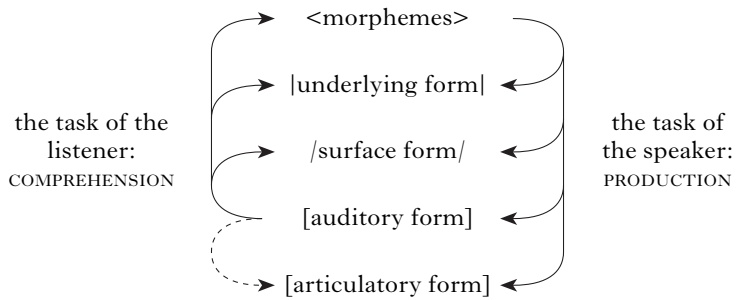


Figure 19
The bigger picture.

this section we discuss the required assumptions and parameters of our model and compare them with those of earlier models.

7.1 General assumption 1: multiple levels of representation

If one wanted to express both phonetic and phonological considerations within OT's usual two-level unidirectional grammar model, one would have to mix detailed phonetic with discrete phonological representations (as in Flemming's 2002 OT version of Dispersion Theory). The reduction to non-teleological underlying mechanisms achieved in the present paper, by contrast, requires us to regard phonetic and phonological representations as separate and as equally important. This is accomplished by the model in Figs 2 and 3, in which the discrete phonological surface form and the detailed phonetic form are distinct representations connected by constraints in a way similar to how faithfulness constraints tend to connect the underlying form and the phonological surface form.

Proposing just one phonological surface form and one phonetic form does not suffice for a full-fledged model of phonology and phonetics. In order for the phonology to be able to handle alternations and metrical structure, we need at least an underlying phonological form as well. And if we want to distinguish more rigorously between the auditory and articulatory aspects of phonetics than we have done here, and if we want to distinguish between form and meaning in the lexicon (Apoussidou 2007: ch. 6), we arrive at the five-level OT (or HG) model in Fig. 19.

The arrows in Fig. 19 have been drawn in parallel to show that although the connections between the representations are local (e.g. cue constraints connect auditory to surface form and faithfulness constraints connect surface to underlying form, but there are no direct connections from auditory to underlying form), the optimisations are performed globally, i.e. by having constraints at different levels interact with each other. In our case of sibilant production, for instance, cue constraints (between auditory and surface form) have to be able to interact with articulatory constraints

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

(at articulatory form); in other cases (Boersma 2007a, b), cue constraints have been shown to interact with faithfulness constraints as well.

The full model of Fig. 19 thus rigorously separates the two discrete phonological representations from the two continuous phonetic representations, and is therefore compatible with many areas of current phonological investigation. At the same time, this model allows the phonology and phonetics to interact through global evaluation, thus reconciling the observation that phonological elements seem to be discrete with the observation that continuous-phonetic considerations seem to influence phonological behaviour.

7.2 General assumption 2: bidirectionality

Beside featuring multiple levels, the model has to be bidirectional, in the sense that the same constraints are used for modelling the speaker and the listener: the prototype effect arises because the constraint ranking that optimises the mapping from phonetic to surface form is reused in production (Boersma 2006).

Within Optimality Theory, bidirectionality has been investigated in some depth by Blutner (2000), Zeevat & Jäger (2002) and Jäger (2003). These proposals (for cases of competition in semantics and pragmatics) try to optimise the listener and the speaker at the same time. Jäger (2003)'s paper is very close to the present one in its general methodology of simulating acquisition and evolution and in its line of thought; for instance, Jäger presents several cases of language types that are both representable in UG and learnable, but unstable over the generations. However, Jäger's Bidirectional Gradual Learning Algorithm relies on a slightly teleological feature of evaluation in production: every candidate form in a production tableau has to be 'hearer-optimal', i.e. if taken as the input to a comprehension tableau (with the same rankings) it should be mapped to a meaning identical to the input of the production tableau.[15] This explicitly listener-oriented evaluation procedure thus militates against ambiguous (i.e. poorly 'dispersed') forms in production, and Jäger relies on it for establishing the diachronic emergence of pragmatic case marking (which enhances the semantic contrast between subject and object). It would be interesting to investigate whether our arguably simpler procedure (optimise comprehension only, then just speak) would be able to handle the complex cases that Jäger discusses.

Outside OT, bidirectionality has been proposed for an exemplar model of phonology. In exemplar theory (Johnson 1997, Lacerda 1997), the listener stores in her memory the actual phonetically detailed tokens ('exemplars') she hears, together with their category labels. Pierrehumbert

[15] A similar criticism can be raised against Boersma's (1998) production model, which relies on a 'control loop', in which every candidate in production is a triplet of articulatory, auditory and surface forms where the auditory form is optimal given the articulatory form and the surface form is optimal given the auditory form.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

(2001) proposes that the listener will subsequently use the same exemplars in production. It has been claimed by both Lacerda (1997: 27) and Pierrehumbert (2001: 143) that a prototype effect similar to the one we derived in §5.5 will arise in listener's goodness ratings (although neither Lacerda nor Pierrehumbert provides the simulations to prove this).[16] Blevins (2004: 285–289) proposes that the shifted 'prototypes' (i.e. the 'best' exemplars in goodness ratings) will be used in production as well (2004: 288), ultimately leading to chain shifts. That this is indeed possible in exemplar theory was proven by the simulations by Wedel (2004: 140–169, 2006: 261–269), which use a device for detecting ambiguous tokens, i.e. tokens whose auditory values lie in the regions of overlap between two categories. By refusing to store the ambiguous tokens as exemplars at all, or by allowing them to be stored in the incorrectly perceived category (rather than in the category intended by the speaker, as happens e.g. in our lexical supervision), Wedel derives the required repulsion of categories.[17] The similarity between Wedel's simulations and ours is that the prototype effect is caused by avoiding regions of overlap (because cue constraints don't move apart in those regions, or because exemplar storage fails in those regions). The difference is that in our model the prototype effect is intrinsic to the bidirectional use of constraints, so that this model, in contrast to Wedel's, makes the empirical prediction that dispersion-like effects must appear in all areas of the grammar, not just in the 'phonetics'. In §8 we speculate on possible applications.

All in all, our model seems to be the one in which the prototype effect in perception and its repulsive effect in production come about most directly and forcefully. Given bidirectional OT with optimisation of comprehension, we would in fact need complex additional machinery if we did *not* want categories to repel each other. The fact that this simplest possible bidirectional OT model already exhibits this effect, an effect that arguably contributes to the success of the species that have it, suggests the idea that biological evolution may well have selected bidirectional constraint competition as a general decision-making mechanism, especially if similar phenomena will turn out to occur in other parts of the grammar (see again §8).

[16] Lacerda did do some computations, but only for the perceptual-magnet effect. Pierrehumbert ran simulations for production, but did not implement the reciprocal inhibition that would have been necessary to illustrate the prototype effect in perception, let alone in production; in fact, Wedel (2004: 195) points out that in Pierrehumbert's simulations cross-category blending in assembling production targets will eventually cause all categories to merge into one. Pierrehumbert's simulations are therefore an instance of the clustering algorithms mentioned in §2.1.6. Following an argument of Wedel (2006: 259–261), we can show that the same is true of Lacerda's (1997) computations.

[17] Interpreting Luce & Pisoni's (1998) experimental results rather freely, Wedel (2006: 264) claims that psycholinguistic evidence supports the existence of such lexical access failures.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

7.3 Assumptions required to go beyond the limitations of the present paper

Our present simple implementation of the model comes with several limitations.

Consider the problem mentioned in §5.4 that it is unrealistic to assume an initial state with fully random prelexical perception together with perfect lexical representations. In a more realistic simulation we will have to take into account the likely fact that the contrast between the categories /s/ and /ʃ/, which has to have been acquired before these categories can be used in lexical entries, has emerged in the learner as the result of prior distributional learning, which should have given a non-trivial initial ranking of the cue constraints. An explicit proposal was provided by Boersma *et al*. (2003). We will also have to take into account the likely fact that lexical representations are acquired in concurrence with prelexical perception. An explicit proposal was provided by Apoussidou (2007). Both refinements will require a comprehensive simulation that is far outside the scope of the present paper.

Another limitation is that our model makes no explicit provisions for diachronic merger. This is because we assumed that even if two categories were very close to each other, as in the initial situation of Fig. 16, the learner was given correct labels by her lexicon. However, a beginning learner does not know beforehand how many sibilants her future language has, and therefore at least has to rely on a stage of distributional learning, in which she establishes the number of categories: if the distributions of two adjacent adult categories overlap too much, these distributions may not form separate peaks and the infant may posit a single category instead of two (e.g. Boersma *et al*. 2003); if she does, a merger has occurred. The comprehensive simulation mentioned in the previous paragraph will take care of this situation. Thus, the present model, combined with the independently needed earlier stage of category creation, accounts for both chain shift and merger.

In the present paper we have limited ourselves to a case with a single auditory dimension. Boersma (2007c) applies the present model to a case with two auditory continua, namely the simulation of an inventory with five vowels within a vowel space that has two auditory dimensions: height (first formant) and backness (second formant). The number of cue constraints scales linearly with the number of continua: if we divide the height and the backness continuum into 100 points each, we require 200 cue constraints per vowel (Boersma & Escudero 2008). Boersma's (2007c) simulations yield the expected dispersion effects for vowels. It is imaginable that for cases with many auditory dimensions, the simulations might interestingly end up in locally optimal inventories, rather than in the globally optimally dispersed inventories found in the present paper; in fact, globally optimal inventories can be defined in Harmonic Grammar and in Maximum Entropy grammars, but not in OT, as simulations involving eternal optimisation by Boersma (2003a) have shown.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Finally, we have limited our discussions to fixed locations of category centres, rather than acknowledging the fact that speakers adapt their auditory spaces to pragmatic circumstances on the fly. Such facts could be included in a way analogous to Boersma's (1998: 208–215) modelling of the dependence of auditory forms on stress and context. Under circumstances of extra articulatory effort (e.g. fast speech), for instance, speakers could indiscriminately raise the rankings of all articulatory constraints by the same amount (Boersma 1998: 275). The auditory values will then become less dispersed immediately (i.e. without learning), just as in the real world. Under circumstances of extra clarity (e.g. addressing a crowd), speakers could raise the rankings of all cue constraints, as well as perhaps those of all faithfulness constraints (Boersma 1998: 275), sensorimotor constraints and lexical constraints. The auditory values are then predicted to become more peripheral immediately (i.e. without learning), just as in the real world. In other words, our model could account for instant adjustments in hyper- and hypospeech (Lindblom 1990a) with an intricacy familiar from the multitude of more discrete constraint-ranking proposals in the OT literature.

7.4 Assumed devices

Beside the general assumptions of multiple representations and bidirectionality mentioned in §§7.1 and 7.2, our model may require some devices more specific to auditory dispersion. In this section we mention these assumptions and compare them with those of the main non-OT account (exemplar theory; see §2.2) and the main previous OT accounts (Flemming, Padgett and Sanders; see §2.3).

Our first assumption is that of the existence of TRANSMISSION NOISE. We think that this is not controversial, and would be included in any explicit model of language transfer. The transmission noise is not represented in the mind or brain of the speaker-listener, but comes for 'free' in the transmission between the speaker's production decision and the listener's auditory-phonetic input.

Our second assumption is that of the greater ARTICULATORY EFFORT associated with peripheral auditory values. This assumption also appears explicitly in Flemming (1995: 46), and would probably be required in Wedel's (2004, 2006) exemplar-based dispersion model if it wanted to account for central tendencies.

Our next assumption involves the way in which the auditory information associated with a category is stored in the brain. One could store this information in a small number of parameters, such as the basic STATISTICS (average and standard distribution) of the spectral mean distribution, or with the help of CATEGORY BOUNDARIES, or as a PROTOTYPE; OT dispersion theory, for instance, seems to require a storage in terms of prototypes, because categories are expressed with auditory values such as 'low F2'. In the present model, none of these parameters is represented, although they could be computed from identification curves such as Fig. 8,

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

distributions such as Fig. 11 or tableaux such as (4). We know of no proposals in the literature that are based on stored statistics, boundaries or prototypes and that come with a learning procedure that exhibits dispersion effects. Another problem with representing such parameters would be that in a full model they would have to be *additional* representations, appearing in between a granular auditory representation, such as the one coming in from the auditory nerve, and the phonological surface level at which the categories are represented; in the present model, the cue constraints *are* the direct connections between the auditory and the phonological level.

An alternative way of storing the auditory information is by using episodic memory. Exemplar theory requires the listener to store every category as a set of multiple EXEMPLARS, each of which records all auditory values associated with a specific auditory event that the listener has heard; our model does not need to store exemplars, because the frequency effects that exemplar theory ascribes to the multiplicity of exemplars tend to derive automatically from the rankings of the many constraints. The most costly method of storing the auditory information would be in terms of joint DISTRIBUTIONS of auditory values. For instance, the vowel space discussed in §7.3 would require 10,000 points to maintain a joint distribution for the two continua; in other words, models that store distributions directly scale poorly (i.e. exponentially) with the number of dimensions. In our model, the distributions are only very indirectly represented in the rankings of the cue constraints, and the model scales well (i.e. linearly) with the number of dimensions (§7.3). We cannot yet assess how the exemplar models of Lacerda (1997) and Pierrehumbert (2001) scale with increasing dimensionality, because they only address single continua.

For the stability of the STANDARD DEVIATION of a category over the generations, our model has to assume nothing: the stability is an automatic consequence of the balance between the entrenchment effect of the articulatory constraints (§5.6) and the transmission noise. By contrast, exemplar theory has to invoke special measures to keep the standard deviation stable, as noted by Pierrehumbert (2001: 149–152): if speakers just randomly choose from among their previously stored exemplars, the transmission noise will soon increase the variation between the exemplars of the category; as a consequence, Pierrehumbert has to propose that speakers use an average of multiple exemplars, and this leads to the required entrenchment. Wedel (2006) calls this averaging 'within-category blending inheritance'; the present model, by contrast, can get by with choosing a single existing auditory value.

Our model does not have to represent or compute AUDITORY DISTANCE. In most exemplar models, listeners need auditory distances to compute the degree to which an existing exemplar is activated (Nosofsky 1988: 701, Kruschke 1992: 23, Johnson 1997: 147, Pierrehumbert 2001: 141; an exception is Lacerda 1997), and when production is included in the model, speakers need auditory distance to compute the contribution of various

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

exemplars in establishing the 'entrenchment' that has to undo the effects of the transmission noise (Pierrehumbert 2001: 149). A very far-reaching representation of auditory distance is needed in OT Dispersion Theory (Flemming 1995) to assess violations of MINDIST. For instance, a language user with categories (or rather, category prototypes) A, B and C has to compute all distances A–B, B–C and A–C, and subsequently to decide which of these is the smallest. In all these models, the computation of auditory distance is regarded as a phenomenon that comes for free; to complete these models, however, an explicit account would have to be given of the underlying mechanism that computes such auditory distances, which seems to require a non-trivial future extension to these models. In our model, by contrast, auditory distance does not have to be computed directly at any point, nor is it represented explicitly at any point; auditory distance *effects* do arise (as in the identification curves of Fig. 8), but they emerge indirectly from the underlying mechanism of acquired constraint ranking, which reflects auditory confusability more directly than auditory distance (see note 13 for some empirical support). In our model, a possible unwanted side effect of not representing auditory distance, namely the learnability of the non-existent or very rare non-contiguous categories, is counteracted by the inherent instability of such categories over the generations (§6.4).[18]

Finally, the present model does not represent any devices explicitly oriented to the goal of improving dispersion, such as the dispersion constraints MINDIST (Flemming 1995), SPACE (Padgett 2003a, b) or $\mathcal{D}_n$-*P* (Sanders 2003). Whether this means that dispersion constraints are superfluous for phonological theory depends on the question of whether they can be used for other things beside evaluating inventories. For instance, dispersion constraints have been used to describe diachronic phonetic enhancement (Sanders 2003: 123). Since phonetic enhancement is also a feature of our model (§§5.9, 6.1), dispersion constraints do not seem to be required for modelling such phenomena. An opposing use of MINDIST has been to describe neutralisation effects: because of the way MINDIST has been formulated (loosely 'small auditory distances between contrasting categories are not allowed'), one way to satisfy it is to neutralise the contrast completely. One such effect, namely unconditional diachronic merger, is handled in our model by the innocent misapprehension that takes place in the infant's distributional learning stage (§7.3). The other effect, namely conditional merger, involves the phonologically more interesting cases of assimilation and positional neutralisation (e.g. Flemming 1995: 119–151). The bidirectional phonology and phonetics model generally accounts for such phenomena by ranking cue constraints over faithfulness constraints (Boersma 2007a: 2035).[19] A rigorous analysis in these terms of

[18] Note that the Gaussian dents that represent local resolution in our model (§5.3) are not a covert trick to measure auditory distance: the prototype effect does not depend on their existence (§7.5).

[19] In Boersma's example, the underlying form |kɛl#ʔazaʁ| can be realised as the phonological-phonetic output pair /kɛlazaʁ/ [kɛlazaʁ], with deletion of the under-

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

all effects ascribed to dispersion constraints in the literature falls outside the scope of the present paper.

7.5 Sensitivity to the parameter settings

The present model requires six parameters: the shape and height of the fixed articulatory effort curve (Figs 10 and 17), the amount of transmission noise (0·8 Erb), the plasticity (0·01), the number of training data (1 million per category), the granularity of the spectral mean continuum (0·1 Erb) and the effective resolution (0·5 Erb).

The results are not qualitatively sensitive to the exact values of these parameters. Raising the effort curve or making it steeper toward the sides will move the emerging category centres further towards each other, whereas raising the transmission noise will move them further apart. Raising the plasticity by a factor of 10 moves the curves of the cue constraints slightly further apart, although this effect can be compensated by reducing the number of training data by a factor of 10. Even reducing the number of training data by 99%, i.e. to 20,000 (thus stopping after the second step in Fig. 7), gives very similar results to those found in the reported simulations: the only difference is that the category centres will turn up somewhat closer together (e.g. by 15% for English). Raising the granularity of the spectral mean continuum by a factor of 10, so that the spacing is 0·01 Erb, does not change any results at all: the results will be identical to the ones reported as long as the spacing is several times smaller than the transmission noise and the effective resolution; this proves that the Gaussian dent method of §5.3 is a correct way to turn a continuum into a set of discrete points. Finally, reducing the effective resolution from 0·5 Erb to zero, effectively making the constraints insensitive to their neighbours, causes fewer tokens to appear in each spectral mean value, hence a reduction of the distance by which the cue constraints will move; but the overall phenomena (prototype effect, entrenchment, stability) do not qualitatively change.

7.6 Dependence on the specific framework of constraint interaction

Our model does not depend on the particular framework in which one models constraint interaction. Although our simulations in §§5 and 6 were done in Stochastic OT, they can equally well be done with Noisy HG (Boersma & Escudero 2008), which is Harmonic Grammar (Smolensky & Legendre 2006) with additive evaluation noise, or with Maximum Entropy models (Goldwater & Johnson 2003). For the simple one-dimensional learning cases in §5.3, Noisy HG is identical to Stochastic

lying glottal stop in a position where it is poorly audible (namely after consonants). This is formalised as the ranking of the cue constraint *[Cʔ̲V]/CʔV/ above the faithfulness constraint Max(ʔ).

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

OT, and the learning algorithm for Noisy HG (Boersma & Escudero 2008) is identical to the Gradual Learning Algorithm for Stochastic OT (the algorithm for Noisy HG is generally identical to Soderstrom *et al.*'s 2006 learning algorithm for non-noisy HG). For our case, the difference between Stochastic OT and Noisy HG therefore lies not in perception, but only in production. In Noisy HG, the height of the articulatory effort curve has no influence on the result; only its shape counts: with the shape of Fig. 10, two-category learners will end up with category centres of 25·6 and 30·4 Erb. For Maximum Entropy grammars, which have a different way of determining winning candidates, tableau (3) still holds: the learning algorithm (Jäger 2007) is generally identical to that of HG and Noisy HG. In Maximum Entropy, the shape of the acquired perception grammar is similar, though not identical, to Fig. 7; however, no entrenchment effect is produced, so that a succession of two-category learners will ultimately end up smoothing out the two categories; a stable situation is reached only if the transmission noise is unrealistically set to zero. We conclude that with all three constraint-based frameworks, the prototype effect is in full operation, but only Stochastic OT and Noisy HG exhibit the equally required entrenchment effect.

7.7 Are phonological features innate or emergent?

In this paper we have used /ʃ/ and /s/ as arbitrary labels that had no preference for any specific positions along the spectral mean continuum. Given this arbitrariness, our model seems slightly more compatible with the viewpoint that phonological features themselves are emergent (Boersma 1998, Blevins 2004, Mielke 2004) than with the viewpoint that phonological features are innate. This is because innate features just seem to be in the way of our learning procedures. For instance, if distributional learning creates three peaks, and therefore three categories, along an auditory continuum, it is easier to give these categories arbitrary labels than to associate them to innate feature values, a procedure that would require an additional mapping device. We realise, though, that the controversial issue of innateness *vs.* emergence is too big for the present paper.

8 Conclusion: the innocent emergence of optimal dispersion

The two assumptions of multiple levels and bidirectionality have explained the origins and stability of auditory dispersion over the generations. It has turned out that if the auditory category centres are too wide apart, the learner's articulatory effect will be greater than her prototype effect, forcing her to shift her production partly towards a smaller, more naturally dispersed inventory, and that if the auditory category centres are too close, the learner's articulatory effect will be smaller than her

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

prototype effect, forcing her to shift her production towards a larger, again more naturally dispersed inventory. Over the generations, every language innocently evolves towards a stable, typologically natural inventory that strikes an optimal balance between articulatory ease and auditory contrast. Everything else being equal, we expect a stable language with two or three sibilants to have the same auditory inventory as English or Polish. In reality, of course, the inventory will be influenced by the rest of the phonological system of the language, because *tout se tient*.

Our findings are relevant for constraint-based phonological theory, because they show that dispersion effects emerge automatically and non-teleologically in the phonology–phonetics interface, and that therefore phonological theory does not have to take them into account at higher levels such as in the phonological surface form (by dispersion constraints) or in the relation between underlying and surface form (by faithfulness constraints).

By deriving dispersion directly from independently needed cue constraints and the independently defended idea of bidirectionality, the results provide support for the idea that humans use constraint ranking or weighting as a decision mechanism (but no support for more linguistically specific concepts of OT such as factorial typology, innate constraints or richness of the base). We have noted that dispersion can be achieved in exemplar theory as well, namely by extending its basic idea with blending inheritance (Pierrehumbert 2001) and lexical decision failures (Wedel 2004, 2006). Both frameworks are *a priori* plausible, because both have their roots in explicit models of the brain: exemplar theory in models of episodic memory, and OT/HG in connectionist models. The empirical difference between the two frameworks as regards dispersion is that the constraint-based model predicts that dispersion effects pervade every level of processing where connecting constraints are used bidirectionally: the bidirectional use of faithfulness constraints (between surface and underlying form in Fig. 19) yields homophony avoidance, the bidirectional use of lexical constraints (between underlying form and morpheme in Fig. 19) yields homonymy avoidance; even in OT semantics, bidirectionality of constraints is predicted to lead to listener-oriented effects in production. By contrast, current exemplar models do not predict such effects, unless more extensions are added. In the end, whether the correct model of human language processing will be similar to a constraint-based model or to an exemplar-based model, or to a synthesis of the two, will depend on future research on overt human behaviour (linguistics and psycholinguistics) as well as on future research on the underlying mechanism (the brain).

REFERENCES

Apoussidou, Diana (2007). *The learnability of metrical phonology*. PhD dissertation, University of Amsterdam.

Avery, Peter, Elan Dresher & Keren Rice (eds.) (2008). *Contrast in phonology : theory, perception, acquisition*. Berlin & New York: Mouton de Gruyter.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Blevins, Juliette (2004). *Evolutionary Phonology : the emergence of sound patterns*. Cambridge: Cambridge University Press.
Blutner, Reinhard (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics* **17**. 189–216.
Boer, Bart de (1999). *Self-organisation in vowel systems*. PhD dissertation, Vrije Universiteit Brussel.
Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **21**. 43–58.
Boersma, Paul (1998). *Functional phonology : formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam.
Boersma, Paul (2003a). The odds of eternal optimization in Optimality Theory. In Holt (2003). 31–65.
Boersma, Paul (2003b). Review of Tesar & Smolensky (2000). *Phonology* **20**. 436–446.
Boersma, Paul (2006). Prototypicality judgements as inverted perception. In Gisbert Fanselow, Caroline Féry, Ralf Vogel & Matthias Schlesewsky (eds.) *Gradience in grammar : generative perspectives*. Oxford: Oxford University Press. 167–184.
Boersma, Paul (2007a). Some listener-oriented accounts of *h*-aspiré in French. *Lingua* **117**. 1989–2054.
Boersma, Paul (2007b). Cue constraints and their interactions in phonological perception and production. Available as ROA-944 from the Rutgers Optimality Archive.
Boersma, Paul (2007c). The emergence of auditory contrast. Paper presented at the 30th GLOW Colloquium, Tromsø.
Boersma, Paul & Paola Escudero (2008). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. In Avery *et al*. (2008). 271–301.
Boersma, Paul, Paola Escudero & Rachel Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. In Solé *et al*. (2003). 1013–1016.
Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* **32**. 45–86.
Boersma, Paul & Joe Pater (2008). Convergence properties of a gradual learning algorithm for Harmonic Grammar. Available as ROA-970 from the Rutgers Optimality Archive.
Boersma, Paul & David Weenink (1992–2008). *Praat : a system for doing phonetics by computer*. http://www.praat.org.
Borgstrøm, Carl H. (1940). *A linguistic survey of the Gaelic dialects of Scotland*. Vol. 1: *The dialects of the Outer Hebrides*. Oslo: Aschehoug.
Bosch, Louis ten (1991). *On the structure of vowel systems : aspects of an extended vowel model using effort and contrast*. PhD dissertation, University of Amsterdam.
Bradlow, Ann R. (1995). A comparative acoustic study of English and Spanish vowels. *JASA* **97**. 1916–1924.
Breen, Gavan & Veronica Dobson (2005). Central Arrernte. *Journal of the International Phonetic Association* **35**. 249–254.
Broersma, Mirjam (2005). Perception of familiar contrasts in unfamiliar positions. *JASA* **117**. 3890–3901.
Carlton, Terence R. (1991). *Introduction to the phonological history of the Slavic languages*. Columbus: Slavica.
Choi, John D. (1991). An acoustic study of Kabardian vowels. *Journal of the International Phonetic Association* **21**. 4–12.
Dart, Sarah N. (1991). *Articulatory and acoustic properties of apical and laminal articulations*. PhD dissertation, University of California, Los Angeles. Distributed as UCLA Working Papers in Phonetics **79**.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Darwin, Charles (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.
Denes, P. (1955). Effect of duration on the perception of voicing. *JASA* **27**. 761–764.
Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology : Human Perception and Performance* **25**. 1568–1578.
Emeneau, M. B. (1944). *Kota texts*. Part 1. Berkeley & Los Angeles: University of California Press.
Escudero, Paola & Paul Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser & Alexander Williams (eds.) *Proceedings of the 25th Annual Penn Linguistics Colloquium*. Philadelphia: University of Pennsylvania. 71–85. Available as ROA-439 from the Rutgers Optimality Archive.
Escudero, Paola & Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* **26**. 551–585.
Fischer, Markus (2005). *A Robbins-Monro type learning algorithm for an entropy maximizing version of stochastic Optimality Theory*. Master's thesis, Humboldt University, Berlin. Available as ROA-767 from the Rutgers Optimality Archive.
Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A* **222**. 309–368.
Flemming, Edward (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles. Published 2002, London & New York: Routledge.
Flemming, Edward (2003). The relationship between coronal place and vowel backness. *Phonology* **20**. 335–373.
Flemming, Edward (2004). Contrast and perceptual distinctiveness. In Hayes *et al*. (2004). 232–276.
Flemming, Edward (2005). Speech perception and phonological contrast. In David B. Pisoni & Robert E. Remez (eds.) *The handbook of speech perception*. Malden, Mass.: Blackwell. 156–181.
Flemming, Edward (2006). The role of distinctiveness constraints in phonology. Ms, MIT.
Forrest, Karen, Gary Weismer, Paul Milenkovic & Ronald N. Dougall (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *JASA* **84**. 115–123.
Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Variation within Optimality Theory : Proceedings of the Stockholm Workshop*. Stockholm: Department of Linguistics, Stockholm University. 111–120.
Gordon, Matthew (1999). *Syllable weight : phonetics, phonology, and typology*. PhD dissertation, University of California, Los Angeles.
Gordon, Matthew, Paul Barthmaier & Kathy Sands (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* **32**. 141–174.
Hamann, Silke (2003). *The phonetics and phonology of retroflexes*. PhD dissertation, University of Utrecht.
Hamann, Silke (2004). Retroflex fricatives in Slavic languages. *Journal of the International Phonetic Association* **34**. 53–67.
Hardcastle, W. J. (1976). *Physiology of speech production : an introduction for speech scientists*. London: Academic Press.
Harris, James W. (1969). *Spanish phonology*. Cambridge, Mass.: MIT Press.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Harrison, Phil (1997). The relative complexity of Catalan vowels and their perceptual correlates. *UCL Working Papers in Linguistics* **9**. 358–402.
Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) (2004). *Phonetically based phonology*. Cambridge: Cambridge University Press.
Hayes, Bruce & Donca Steriade (2004). Introduction: the phonetic bases of phonological markedness. In Hayes *et al*. (2004). 1–33.
Heffner, Roe-Merrill S. (1937). Notes on the length of vowels. *American Speech* **12**. 128–134.
Hirose, Hajime & Thomas Gay (1972). The activity of the intrinsic laryngeal muscles in voicing control. *Phonetica* **25**. 140–164.
Holt, D. Eric (ed.) (2003). *Optimality Theory and language change*. Dordrecht: Kluwer.
House, Arthur S. & Grant Fairbanks (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *JASA* **25**. 105–113.
Hoyt, Donald F. & C. Richard Taylor (1981). Gait and the energetics of locomotion in horses. *Nature* **292**. 239–240.
Hutters, B. (1985). Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* **42**. 1–24.
Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene & Emmanuel Dupoux (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience* **23**. 9541–9546.
Jäger, Gerhard (2003). Learning constraint sub-hierarchies: the Bidirectional Gradual Learning Algorithm. In Henk Zeevat & Reinhard Blutner (eds.) *Optimality Theory and pragmatics*. Basingstoke: Palgrave Macmillan. 251–287.
Jäger, Gerhard (2007). Maximum entropy models and Stochastic Optimality Theory. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Chris Manning (eds.) *Architectures, rules, and preferences : variations on themes by Joan W. Bresnan*. Stanford: CSLI. 467–479.
Jakobson, Roman (1941). *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Lundequist.
Jassem, Wiktor (2003). Polish. *Journal of the International Phonetic Association* **33**. 103–107.
Johnson, Keith (1997). Speech perception without speaker normalization: an exemplar model. In Keith Johnson & John W. Mullennix (eds.) *Talker variability in speech processing*. San Diego: Academic Press. 145–165.
Johnson, Keith, Edward Flemming & Richard Wright (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Lg* **69**. 505–528.
Jones, Mark (2001). The historical development of retroflex fricatives in Polish: markedness, functionality, phonology and phonetics. Ms, Trinity College, Cambridge.
Jongman, Allard, Ratree Wayland & Serena Wong (2000). Acoustic characteristics of English fricatives. *JASA* **108**. 1252–1263.
Keating, Patricia A. (1979). *A phonetic study of a voicing contrast in Polish*. PhD dissertation, Brown University.
Keating, Patricia A. (1985). Universal phonetics and the organization of grammars. In Victoria A. Fromkin (ed.) *Phonetic linguistics : essays in honor of Peter Ladefoged*. Orlando: Academic Press. 115–132.
Keating, Patricia A., Wendy Linker & Marie Huffman (1983). Patterns in allophone distribution for voiced and voiceless stops. *JPh* **11**. 277–290.
Kirchner, Robert (1998). *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles. Published 2001, New York & London: Routledge.
Kochetov, Alexei (2008). Self-organization through misperception: secondary articulation and vowel contrasts in language inventories. In Avery *et al*. (2008). 193–216.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Kruschke, John K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review* **99**. 22–44.
Lacerda, Francisco (1997). Distributed memory representations generate the perceptual-magnet effect. Ms, Stockholm University.
Ladefoged, Peter (2001). *Vowels and consonants*. Malden, Mass. & Oxford: Blackwell.
Ladefoged, Peter (2003). *Phonetic data analysis: an introduction to fieldwork and instrumental techniques*. Malden, Mass. & Oxford: Blackwell.
Ladefoged, Peter, Jenny Ladefoged, Alice Turk, Kevin Hind & St. John Skilton (1997). Phonetic structures of Scottish Gaelic. *UCLA Working Papers in Phonetics* **95**. 114–153.
Ladefoged, Peter & Ian Maddieson (1996). *The sounds of the world's languages*. Oxford: Blackwell.
Ladefoged, Peter & Zongji Wu (1984). Places of articulation: an investigation of Pekingese fricatives and affricates. *JPh* **12**. 267–278.
Liljencrants, Johan & Björn Lindblom (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Lg* **48**. 839–862.
Lindblom, Björn (1986). Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology*. Orlando: Academic Press. 13–44.
Lindblom, Björn (1990a). Explaining phonetic variation: a sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer. 403–439.
Lindblom, Björn (1990b). Models of phonetic variation and selection. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* **11**. 65–100.
Luce, Paul A. & Jan Charles-Luce (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *JASA* **78**. 1949–1957.
Luce, Paul A. & David B. Pisoni (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing* **19**. 1–36.
McCarthy, John J. (2002). *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.
Maddieson, Ian (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
Maddieson, Ian (1987). The Margi vowel system and labiocoronals. *Studies in African Linguistics* **18**. 327–355.
Manuel, Sharon Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *JASA* **88**. 1286–1298.
Mees, Inger & Beverley Collins (1982). A phonetic description of the consonant system of Standard Dutch (ABN). *Journal of the International Phonetic Association* **12**. 2–12.
Mielke, Jeff (2004). *The emergence of distinctive features*. PhD dissertation, Ohio State University.
Moore, Brian C.J. & Brian R. Glasberg (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *JASA* **74**. 750–753.
Morrison, Geoffrey Stewart (2002). *Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels*. MA dissertation, Simon Fraser University, Burnaby.
Navarro Tomás, T. (1932). *Manual de pronunciación española*. 4th edn. Madrid: Centro de Estudios Históricos.
Nosofsky, Robert M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**. 700–708.
Nowak, Paweł M. (2006). The role of vowel transitions and frication noise in the perception of Polish sibilants. *JPh* **34**. 139–152.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Ohala, John J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (eds.) *Papers from the parasession on language and behavior*. Chicago: Chicago Linguistic Society. 178–203.
Oudeyer, Pierre-Yves (2006). *Self-organization in the evolution of speech*. Oxford: Oxford University Press.
Padgett, Jaye (2001). Contrast dispersion and Russian palatalization. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*. San Diego: Academic Press. 187–218.
Padgett, Jaye (2003a). The emergence of contrastive palatalization in Russian. In Holt (2003). 307–335.
Padgett, Jaye (2003b). Contrast and post-velar fronting in Russian. *NLLT* **21**. 39–87.
Padgett, Jaye (2004). Russian vowel reduction and Dispersion Theory. *Phonological Studies* **7**. 81–96.
Padgett, Jaye & Marzena Zygis (2003). The evolution of sibilants in Polish and Russian. *ZAS Working Papers in Linguistics* **32**. 155–174.
Pater, Joe (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In René Kager, Joe Pater & Wim Zonneveld (eds.) *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 219–244.
Pater, Joe (2008). Gradual learning and convergence. *Linguistic Inquiry* **39**. 334–345.
Pierrehumbert, Janet (2001). Exemplar dynamics: word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: Benjamins. 137–157.
Polivanov, E. D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* **4**. 79–96. Translated as 'The subjective nature of the perceptions of language sounds' in E. D. Polivanov (1974). *Selected works : articles on general linguistics*. The Hague: Mouton. 223–237.
Prince, Alan & Paul Smolensky (1993). *Optimality Theory : constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.
Puppel, Stanisław, Jadwiga Nawrocka-Fisiak & Halina Krassowska (1977). *A handbook of Polish pronunciation for English learners*. Warsaw: Państwowe Wydawnictwo Naukowe.
Raphael, Lawrence J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *JASA* **51**. 1296–1303.
Repp, Bruno H. (1981). Two strategies in fricative discrimination. *Perception and Psychophysics* **30**. 217–227.
Ringen, Catherine & Pétur Helgason (2004). Distinctive [voice] does not imply regressive assimilation: evidence from Swedish. *International Journal of English Studies* **4.2**. 53–71.
Romani, Gian Luca, Samuel J. Williamson & Lloyd Kaufman (1982). Tonotopic organization of the human auditory cortex. *Science* **216**. 1339–1340.
Rospond, Stanisław (1971). *Gramatyka historyczna języka polskiego*. Warsaw: Państwowe Wydawnictwo Naukowe.
Sanders, Nathan (2003). *Opacity and sound change in the Polish lexicon*. PhD dissertation, University of California, Santa Cruz.
Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée & Christian Abry (1997). The Dispersion-Focalization Theory of vowel systems. *JPh* **25**. 255–286.
Shalev, Michael, Peter Ladefoged & Peri Bhaskararao (1993). Phonetics of Toda. *UCLA Working Papers in Phonetics* **84**. 89–125. Also published (1994) in *PILC Journal of Dravidic Studies* **4**. 21–56.
Smolensky, Paul (1996). On the comprehension/production dilemma in child language. *LI* **27**. 720–731.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press

Smolensky, Paul & Géraldine Legendre (2006). *The harmonic mind : from neural computation to optimality-theoretic grammar*. 2 vols. Cambridge, Mass.: MIT Press.
Soderstrom, Melanie, Donald Mathis & Paul Smolensky (2006). Abstract genomic encoding of Universal Grammar in Optimality Theory. In Smolensky & Legendre (2006: vol. 2). 403–471.
Solé, M. J., D. Recasens & J. Romero (eds.) (2003). *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions.
Stieber, Zdzisław (1952). *Rozwój fonologiczny języka polskiego*. Warsaw: Państwowe Wydawnictwo Naukowe. Translated by E. Schwartz (1968) as *The phonological development of Polish*. Ann Arbor: University of Michigan.
Stone, Maureen, Alice Faber, Lawrence J. Raphael & Tomas H. Shawker (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *JPh* **20**. 253–270.
Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes & Annabel Greenhill (eds.) *Proceedings of the 21st Annual Boston University Conference on Language Development*. Somerville, Mass.: Cascadilla. 615–626.
Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *LI* **29**. 229–268.
Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.
Tingsabadh, M. R. Kalaya & Arthur S. Abramson (1999). Thai. In *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press. 147–150.
Toda, Martine (2005). Effect of palate shape on the spectral characteristics of coronal fricatives. Paper presented at the *Conference on Turbulences*, ZAS Berlin.
Wedel, Andrew B. (2004). *Self-organization and categorical behavior in phonology*. PhD dissertation, University of California, Santa Cruz.
Wedel, Andrew B. (2006). Exemplar models, evolution and language change. *The Linguistic Review* **23**. 247–274.
Wedel, Andrew B. (2007). Feedback and regularity in the lexicon. *Phonology* **24**. 147–185.
Zeevat, Henk & Gerhard Jäger (2002). A reinterpretation of syntactic alignment. In Dick de Jongh, Marie Nilsenová & Henk Zeevat (eds.) *Proceedings of the 4th International Tbilisi Symposium on Language, Logic and Computation*. University of Amsterdam.
Zygis, Marzena (2003). The role of perception in Slavic sibilant systems. In Peter Kosta, Joanna Błaszczak, Jens Frasek, Ljudmila Geist & Marzena Zygis (eds.) *Investigations into formal Slavic linguistics : contributions of the 4th European Conference on Formal Description of Slavic Languages*. Frankfurt: Peter Lang. 137–153.
Zygis, Marzena & Silke Hamann (2003). Perceptual and acoustic cues of Polish coronal fricatives. In Solé *et al*. (2003). 395–398.

https://doi.org/10.1017/S0952675708001474 Published online by Cambridge University Press