OCP 3, Budapest, January 17, 2006

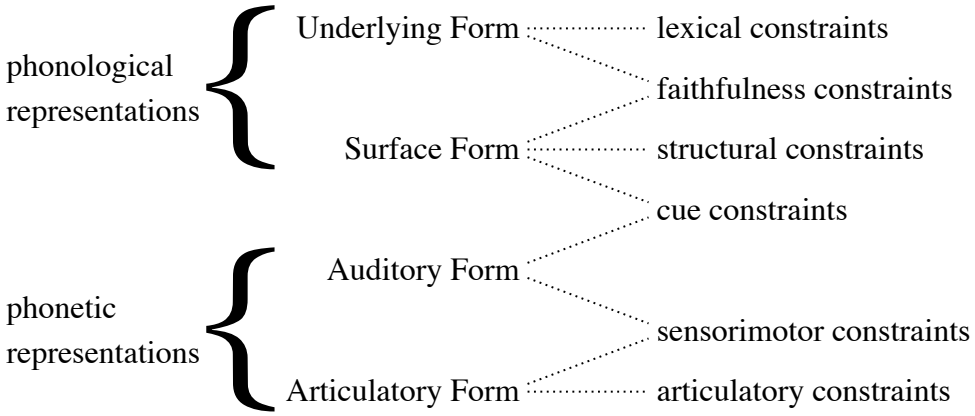
Sibilant inventories in bidirectional phonology and phonetics

Paul Boersma
Universiteit van Amsterdam
paul.boersma@uva.nl

Silke Hamann
ZAS Berlin
silke@zas.gwz-berlin.de

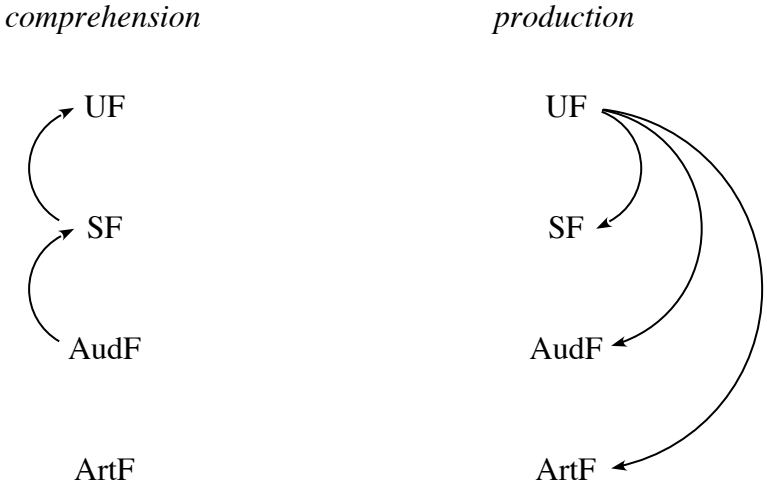
1. Bidirectional phonology and phonetics

1a. A single grammar for phonology and phonetics: four representations



Where is phonetic detail? There are two *discrete economical* representations (underlying and surface form) and two *continuous rich* representations (auditory and articulatory form).

1b. A single grammar for two directions of processing



Comprehension: this process consists of two sequential modules (McQueen & Cutler 1997): the listener first converts a given auditory form into a phonological surface form (*prelexical perception*), which she then uses to find an underlying form in the lexicon (*word recognition*).

Production: this process is parallel (Boersma 2005a): the speaker converts a given UF into an optimal triplet { articulatory form, auditory form, surface form } that is determined by the interaction of faithfulness, structural, cue, sensorimotor, and articulatory constraints.

1c. Constraints (from bottom to top)

Articulatory constraints evaluate articulatory effort in phonetic implementation. They may have a fixed and language-independent ranking (Kirchner 1998) or a ranking that is influenced by language-specific learning (Boersma 1998).

Sensorimotor constraints express the speaker's knowledge of the relation between sound and muscle movements. We propose that the constraints themselves express arbitrary relations and that language-specific learning moves the relevant ones to the bottom of the hierarchy. For instance, the sensorimotor constraint "*[fronted tongue]$_{ArtF}$ [low F2]$_{AudF}$" will be ranked high, while "*[fronted tongue]$_{ArtF}$ [high F2]$_{AudF}$" will be ranked low.

Cue constraints evaluate language-specific cue integration in perception (Escudero & Boersma 2004). For instance, the duration of the preceding vowel is a major cue to obstruent voicing in English but not in most other languages. Hence, the cue constraint "*[long vowel duration]$_{AudF}$ /obs,–voice/$_{SF}$" is ranked high in English but low elsewhere.

Structural constraints evaluate language-specific restrictions on produced structure (Prince & Smolensky 1993: e.g. NOCODA) as well as on perceived structure (Tesar 1997 and Tesar & Smolensky 2000: 'robust interpretive parsing' of metrical structure; e.g. ALLFEETLEFT).

Faithfulness constraints evaluate the relation between the two phonological forms in production (McCarthy & Prince 1995) as well as in recognition (Boersma 2001), e.g. MAX.

Lexical constraints (not considered here at all) militate against recognizing certain lexical items and therefore interact with faithfulness in recognition (Boersma 2001). Their ranking is sensitive to semantic context.

Interaction of structural and cue constraints. These two families interact in perception. For instance, Japanese listeners perceive [ebzo]$_{Aud}$ as /e.bu.zo/$_{SF}$ (Polivanov 1931, Dupoux et al. 1999), because perceiving it as */eb.zo/$_{SF}$ would violate the structural constraint CODACOND. This constraint must outrank the cue constraint *[ ]$_{Aud}$ /u/$_{SF}$ that militates against hallucinating an /u/$_{SF}$ for which there are no direct auditory cues. Such interactions constitute one of the main reasons to regard perception as phonological, and hence to formulate cue constraints as Optimality-Theoretic.

Interaction of faithfulness and cue constraints. Cue constraints are not only used in perception, but in production as well (Boersma 2005ab). Since production is parallel, cue constraints interact with faithfulness. Thus, if both faithfulness and cue constraints are high-ranked, English speakers must lengthen their vowels before a voiced obstruent; if they don't, they would violate either faithfulness (by assuming that the surface form is /–voice/$_{SF}$) or the cue constraint (by assuming that the surface form is /+voice/$_{SF}$). Such interactions yield *licensing by cue* effects without the need to rank faithfulness constraints themselves by cues (Steriade 1995) or probability (Boersma & Hamann 2004).

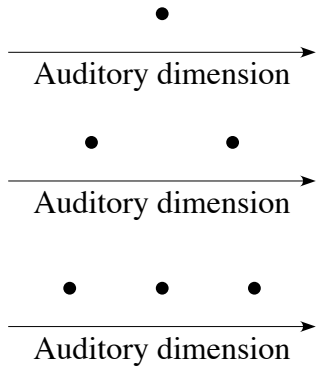
Markedness constraints are superfluous. Markedness effects (both articulatory ease and auditory distinctiveness) arise synchronically as well as diachronically (as we will see) as a result of the workings of articulatory and cue constraints.

2. Perceptual dispersion effects

2a. The dispersion principle before OT

Phonemes in a segmental inventory usually show a perceptual dispersion effect: they are located at equal distances within the auditory space to be perceptually maximally distinct (Liljencrants & Lindblom 1972, Lindblom's 1986 'Theory of adaptive dispersion').

A secondary dispersion effect is that for larger inventories, the auditory space enlarges, but the distance between the segments decreases (phonemic notations, as below, do not always reflect this).



	F2:	*VOT:*	*F2 transition:*
Auditory dimension	/ɨ/ (Margi)	/p/ (C. Arrernte)	/n/ (Dutch)
Auditory dimension	/i u/ (Spanish)	/b p^h/ (Swedish)	/n^j $n^ɣ$/ (Russian)
Auditory dimension	/i ɨ u/ (Polish)	/b p p^h/ (Thai)	/n^j n $n^ɣ$/ (Sc. Gaelic)

2b. Prince & Smolensky (1993): faithfulness constraints

The original OT proposal by Prince & Smolensky handled inventories by the device of *Richness of the Base* and constraint interaction. In their proposal, 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.

Problem. Prince & Smolensky's approach cannot account for dispersion effects, where marked segments appear without the unmarked counterpart.

2c. Flemming (1995): dispersion constraints

Flemming (1995) translated the dispersion idea into OT by introducing MINDIST constraints. These dispersion constraints work very well in formalizing the dispersion idea.

Problems. First, dispersion constraints do nothing else beside explaining inventories. Second, these constraints evaluate multiple inputs at a time and are therefore hard to reconcile with the single-input constraints introduced by Prince & Smolensky.

2d. Sanders (2003): faithfulness plus dispersion constraints

Sanders (2003) uses both MINDIST constraints and faithfulness, avoiding the single-input problem of Flemming's approach.

Problem. One problem persists: the dispersion constraints do nothing else beside explaining inventories.

2e. Boersma & Hamann (today): neither faithfulness nor dispersion constraints

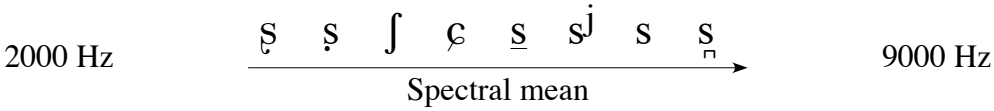
We claim that dispersion effects instead automatically arise within a couple of generations as the automatic result of independently needed constraints:

- **cue constraints** (which are independently needed to model language-specific perception)
- **articulatory constraints** (which are independently needed to model articulatory effort in phonetic implementation)

3. Sibilant inventories: dispersion along the spectral mean

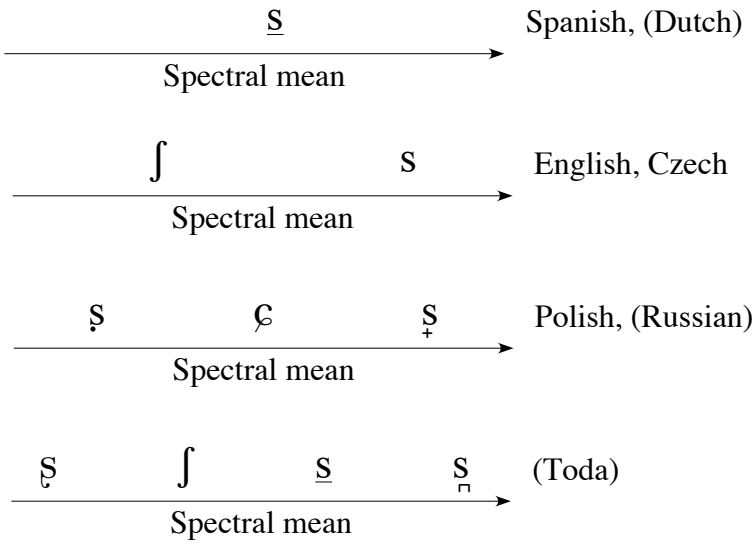
3a. The major cue for sibilant articulator and place

The sibilants in a language can often be ordered along a one-dimensional auditory continuum, namely the *spectral centre of gravity* or *spectral mean* (e.g. Forrest et al. 1988, Gordon et al. 2002). Articulatorily, the spectral mean correlates with the frontness of the tongue and with the frontness of the place of articulation.



Simplification: we pretend in this paper that this is the only relevant perceptual cue. In reality, other cues for sibilant articulator and place include spectral peaks (e.g. Jongman et al. 2000) and vowel transitions (e.g. Nowak, to appear).

3b. Sibilant inventories tend to be dispersed



3c. Modelling sibilant dispersion in OT

The Polish sibilant inventory has been formalized with dispersion constraints by Padgett & Zygis (2003). In the following we will show that it can be done without.

4. Sibilant evolution in bidirectional phonology and phonetics

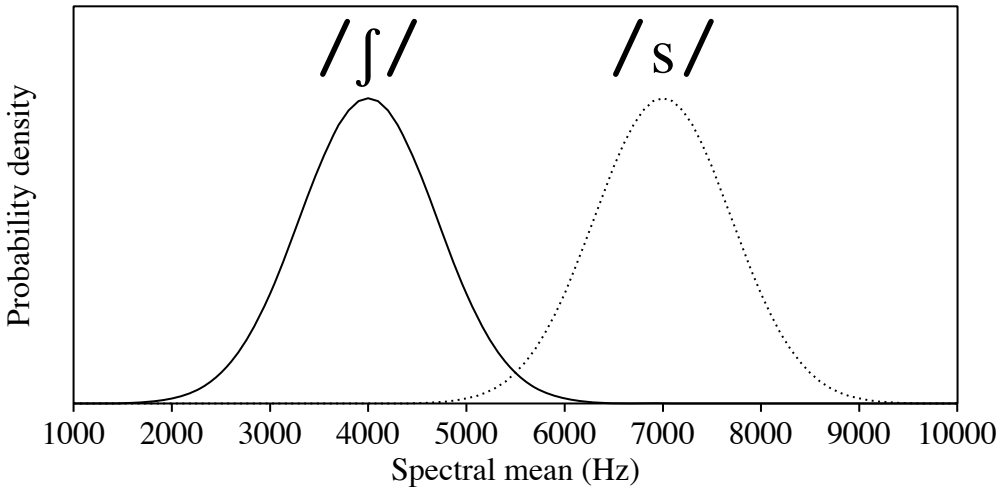
4a. The English sibilant inventory

English has two sibilants:

- /s/ (average spectral mean 7000 Hz)
- /ʃ/ (average spectral mean 4000 Hz)

4b. The English listener's auditory environment distribution

Given the averages of 4a, and making an informed guess about the variation within and between speakers and about noise caused by the transmission (muscles, air, ear), the English listener will hear the following spectral-mean distributions for the two sibilants:

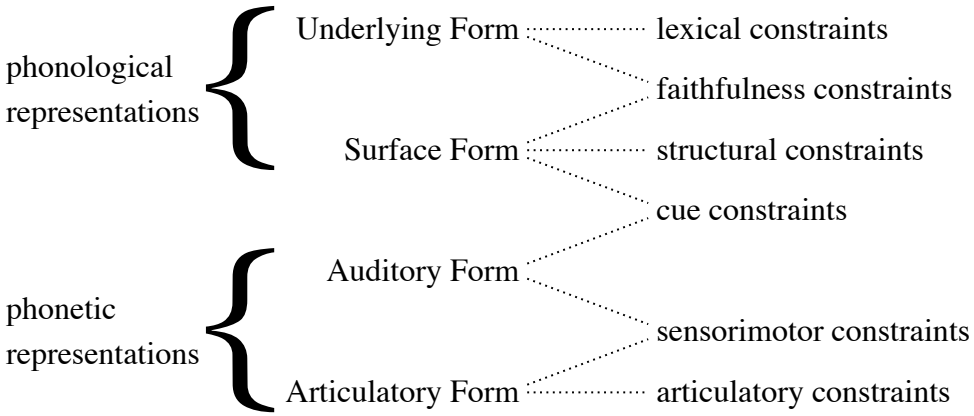


4c. What is the optimal way for an English listener to perceive the spectral mean?

In order to prepare best for lexical access, prelexical perception should make the fewest 'mistakes' in classifying incoming auditory tokens. For the distribution of 4b, this means that an optimal listener has to classify spectral-mean tokens below $[5500\ \text{Hz}]_{\text{AudF}}$ as $/ʃ/_{\text{SF}}$, and tokens above $[5500\ \text{Hz}]_{\text{Aud}}$ as $/s/_{\text{SF}}$.

4d. How does the listener manage to perceive these in a way optimal for English?

Our answer:



4e. Cue constraints for classifying spectral mean in English

We propose (with Escudero & Boersma 2004) that cue constraints express all possible arbitrary relationships between AudF and SF. For English, this means that cue constraints for the spectral-mean-to-sibilant mapping connect any possible spectral-mean frequency between [1000 Hz] and [10000 Hz] to either sibilant category. Examples are:

*[1000 Hz]/s/	*[1000 Hz]/ʃ/
*[1100 Hz]/s/	*[1100 Hz]/ʃ/
...	...
*[9900 Hz]/s/	*[9900 Hz]/ʃ/
*[10000 Hz]/s/	*[10000 Hz]/ʃ/

For perception, these constraints can be read as "a spectral mean of [1000 Hz] should not be perceived as /s/" and so on.

Simplification: we discretize the spectral-mean range into 91 steps of 100 Hz. In reality, the resolution is based on hair cells and auditory nerve fibers.

4f. A perception tableau for classifying spectral mean in English

In 4b and 4c we see that a good listener of English should perceive $[4700\ \text{Hz}]_{\text{Aud}}$ as $/ʃ/_{\text{SF}}$. The following tableau shows a possible ranking that achieves this:

$[4700\ \text{Hz}]_{\text{Aud}}$	*[4600]/s/	*[4700]/s/	*[4800]/s/	*[4800]/ʃ/	*[4700]/ʃ/	*[4600]/ʃ/
$/s/_{\text{SF}}$		*!				
☞ $/ʃ/_{\text{SF}}$					*	

In itself, proposing such a ranking of 182 constraints is rather unconvincing. To increase plausibility, we will *explain* how a learner derives the ranking with the help of a simple learning procedure and algorithm.

4g. The learning procedure: lexicon-driven learning of perception

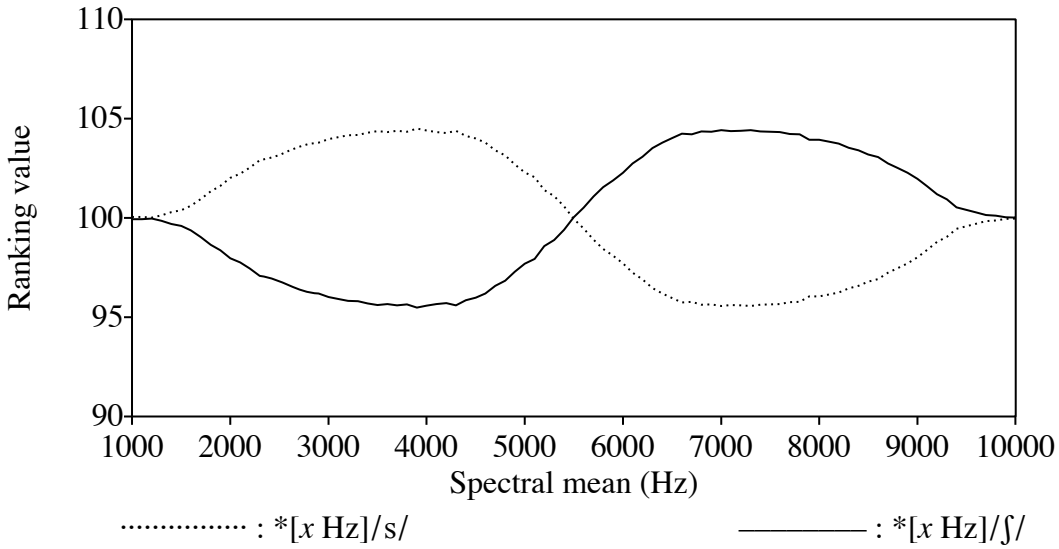
Initial state. We start modelling the acquisition process at the point where the learner already has correct lexical representations (i.e. she knows which lexical items have $|s|_{\text{UF}}$ and which have $|ʃ|_{\text{UF}}$) but prelexical perception is still far from adult-like.

Subsequent development. The child will hear auditory spectral-mean tokens (drawn from the distribution in 4b) that her lexicon will subsequently label correctly as $|s|_{\text{UF}}$ or $|ʃ|_{\text{UF}}$. This can lead to *lexicon-driven perceptual learning* with the *Gradual Learning Algorithm* (Boersma 1997). For instance, the child in the tableau below receives a spectral mean of [4700 Hz] that her current non-optimal grammar perceives as /s/ (☞). If her lexicon subsequently tells her that she should have perceived /ʃ/ (√) instead (perhaps because the recognized word was *sheep*), two cue constraints will be reranked slightly in the direction of the arrows.

$[4700\ \text{Hz}]_{\text{Aud}}$	*[4600]/s/	*[4800]/ʃ/	*[4700]/ʃ/	*[4600]/ʃ/	*[4700]/s/	*[4800]/s/
☞ $/s/_{\text{SF}}$					←*	
√ $/ʃ/_{\text{SF}}$			*!→			

4h. A computer simulation of English perception

A virtual learner has 182 cue constraints, all ranked initially at 100.0 (which is another simplication). The learner then hears 2 million spectral-mean tokens randomly drawn from the distributions in 4b (with an equal probability of 50% for each sibilant). Each token is labelled correctly as /s/ or /ʃ/ by the lexicon, so that the learning algorithm of 4g can do its work. With an *evaluation noise* (per-tableau random variation in ranking) of 2.0, and a *plasticity* (reranking step) of 0.01, the following perception grammar results:



At any specific spectral-mean value, the lowest curve determines which category is perceived most often. For instance, we see that *[4700 Hz]/ʃ/ lies below *[4700 Hz]/s/, so that [4700 Hz] will be perceived most often as /ʃ/ (not always, because of the evaluation noise).

4i. English production

In a *bidirectional* model of phonology and phonetics, the rankings in the perception grammar must also be used in production. The following tableau shows how an underlying $|s|_{UF}$ will be produced if faithfulness is high-ranked:

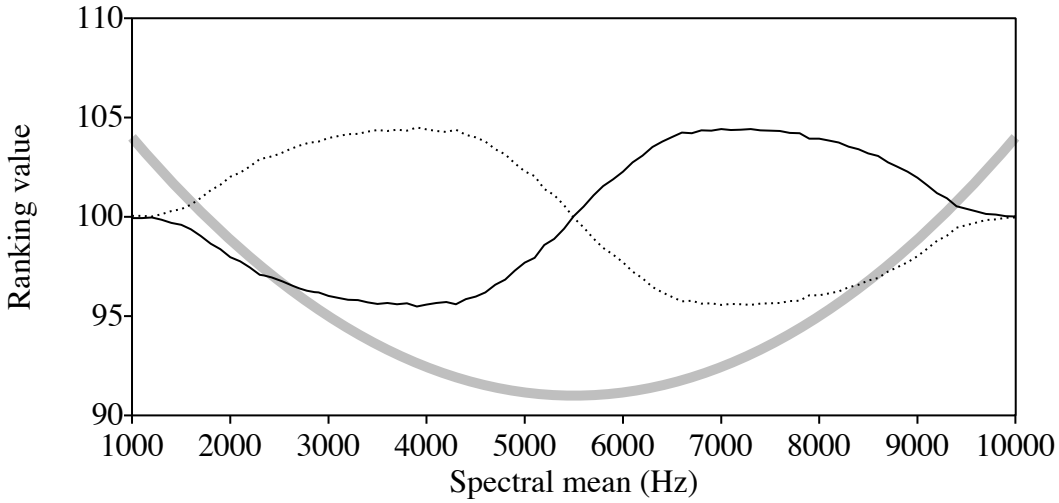
$\|s\|_{UF}$	FAITH	*[7200] /s/	*[7000] /s/	*[7100] /s/	$*[7200]_{ArtF}$	$*[7100]_{ArtF}$	$*[7000]_{ArtF}$
$/s/_{SF}\,[7000\text{ Hz}]_{AudF}$			*!				*
☞ $/s/_{SF}\,[7100\text{ Hz}]_{AudF}$				*		*	
$/s/_{SF}\,[7200\text{ Hz}]_{AudF}$		*!			*		
$/ʃ/_{SF}\,[4000\text{ Hz}]_{AudF}$	*!						

The candidate $/s/_{SF}\,[7100\text{ Hz}]_{AudF}$ wins because the curve of the cue constraints for /s/ in 4h is lowest at 7100 Hz. The constraint $*[7200]_{Art}$ is an articulatory constraint whose ranking reflects the articulatory effort associated with producing a spectral mean of 7200 Hz.

Simplification. We assume that the sensorimotor constraints are ranked perfectly, so that the articulation that goes with a specific spectral mean can be assumed fixed. As a result, the candidates can be doublets { SF, AudF/ArtF } rather than triplets { SF, AudF, ArtF }.

4j. A balance between the prototype effect and articulatory constraints

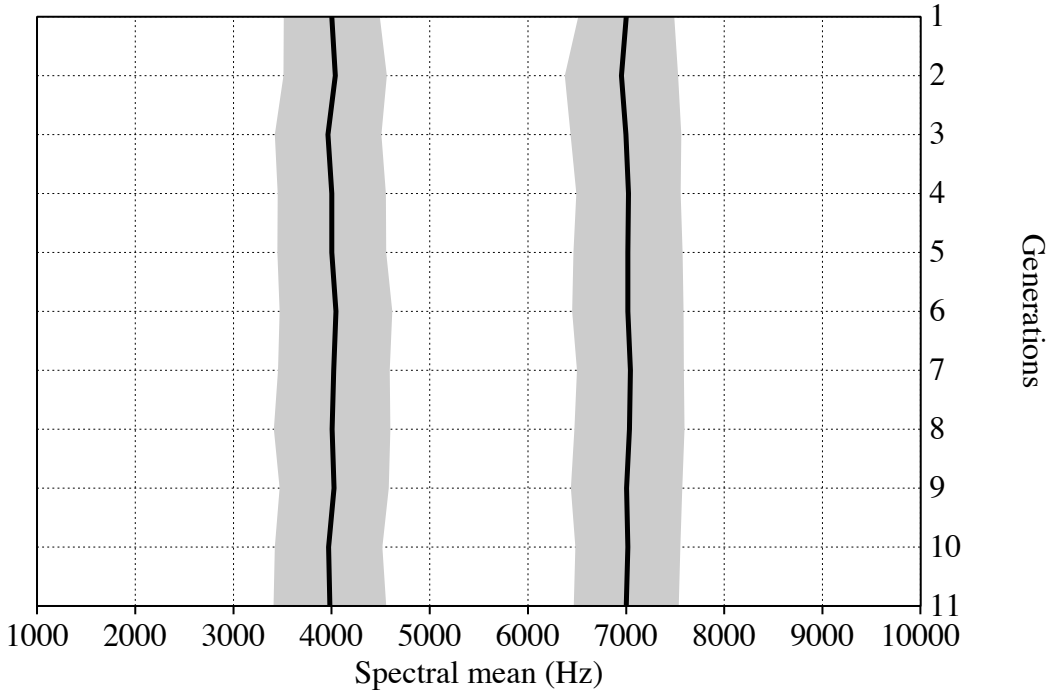
If the cue constraints of 4h are used for producing an /s/, its resulting average spectral mean will be around 7100 Hz (this is where the curve is lowest), i.e. the spectral mean has shifted by 100 Hz in one generation. This is the *prototype effect* (Boersma 2005b). An unbridled dispersion over the generations can only be prevented by articulatory constraints (drawn here in thick grey), which express the fact that peripheral auditory values are hardest to produce:



With this grammar, the English learner will produce average spectral means of 4000 and 7000 Hz, just like her parents (the tableau in 4i looks as though this is incorrect, but the evaluation noise will cause *[7100]$_{\text{ArtF}}$ to outrank *[7100]$_{\text{AudF}}$/s/$_{\text{SF}}$ in a small fraction of the evaluations).

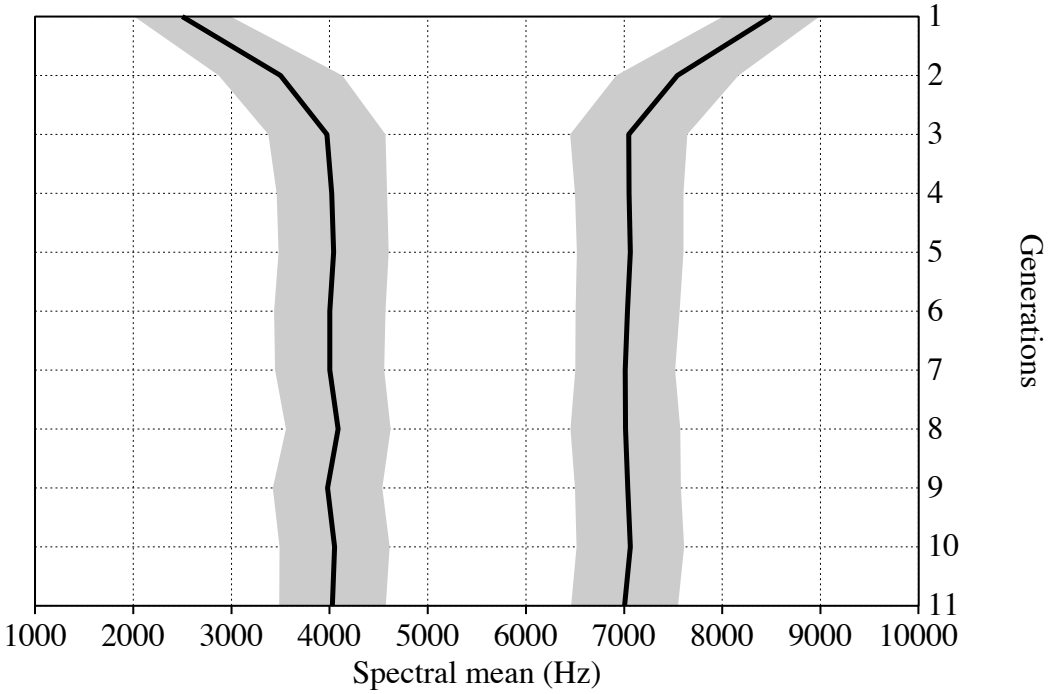
4k. English: a stable sibilant system

The English spectral mean values of 4000 and 7000 Hz will stay constant over the generations (the simulation had 100000 data per generation, and a plasticity of 0.1; the black lines in the figure are single-speaker averages, the grey areas are within-speaker standard deviations):



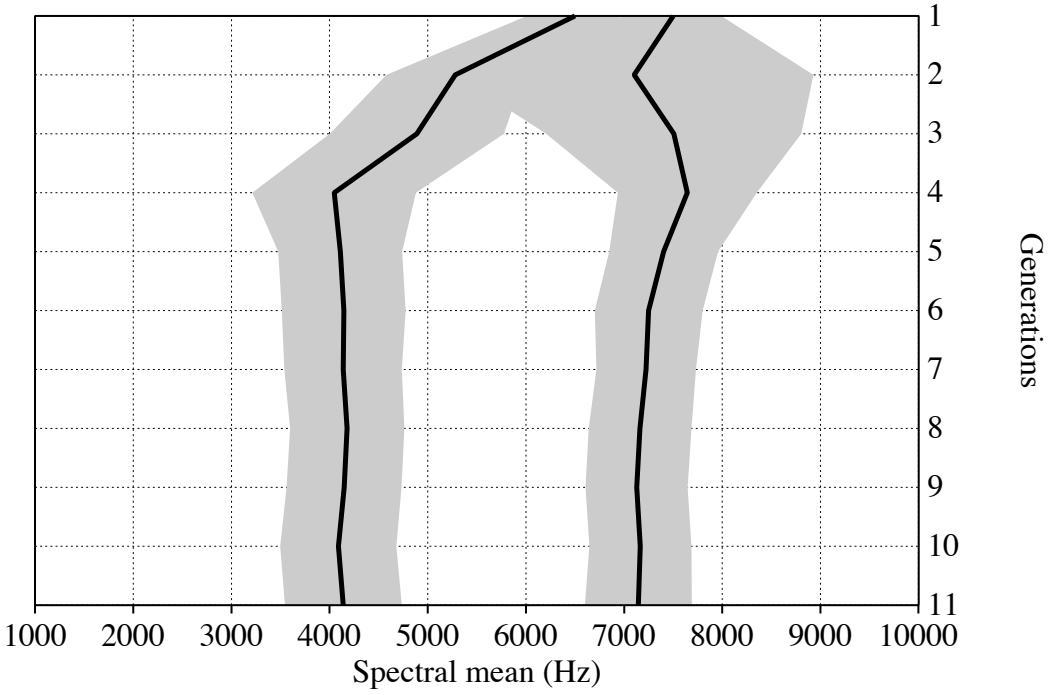
4l. 'Exaggerated English': learnable but unstable

If a language starts out with a much more extreme two-sibilant contrast than English, i.e. [s̻] versus [s̺] (spectral means of 2500 and 8500 Hz), the second generation will more or less learn the oversized range, but the third generation will already have shifted the system towards an unmarked articulatorily-perceptually balanced [ʃ] and [s]:



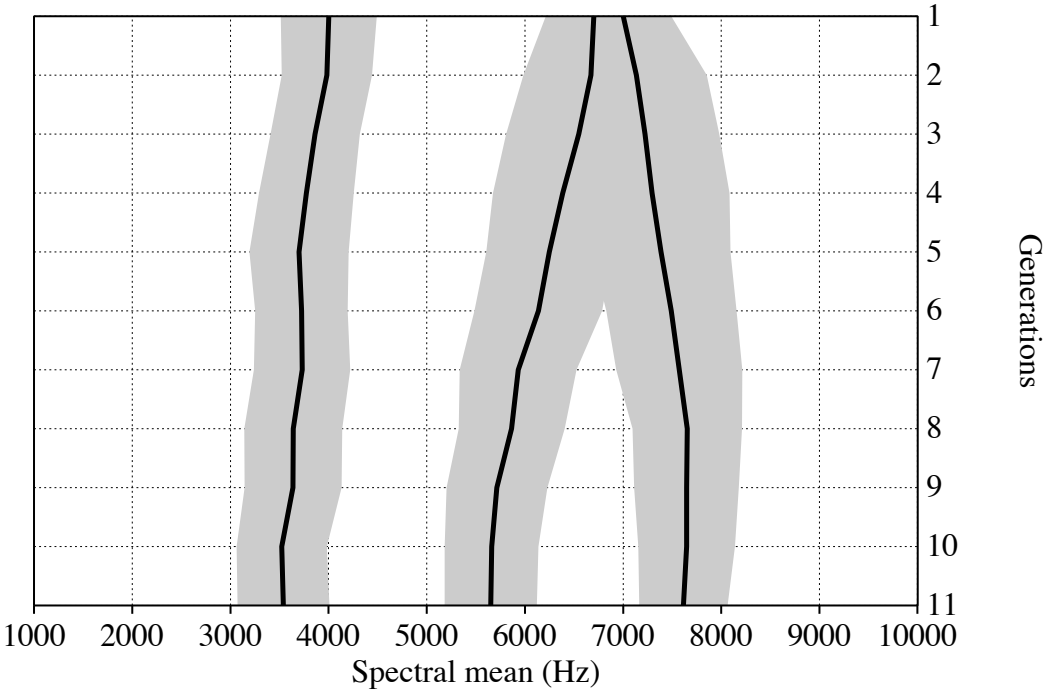
4m. 'Confusing and skewed English': learnable but unstable

If the inventory is skewed (both sibilants have a high spectral mean) and confusing (their difference is only 1000 Hz), three generations will still suffice to render it English-like:

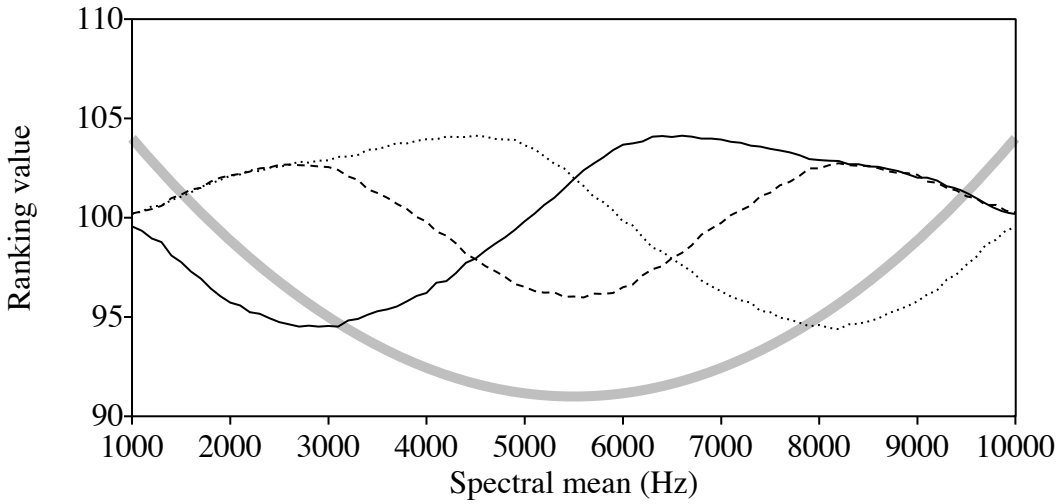


4n. The Polish three-sibilant system

Polish used to have a /ʃ sʲ s/ system. If the dispersion principle is correct, and if the spectral mean is the only auditory cue that distinguishes these three sibilants, such a system cannot be stable. The following simulation shows what happens if we start out with a language whose sibilants have spectral means of 4000, 6700 and 7000 Hz (in order to make our point, we have taken these last two closer together than they probably actually were in medieval Polish).



The first striking phenomenon that the figure shows is that the palatalized alveolar lowers its spectral mean beyond 6000 Hz, 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 first shift causes the postalveolar to shift down towards /ʂ/. This is reported to have happened in real Polish in the 16th century (Rospond 1971). This simulation thus explains the present Polish sibilant system, which has spectral means not far removed from the 3500, 5500, and 7500 Hz found here (Zygis & Hamann 2003). With these three spectral means in her auditory environment, a Polish learner will acquire the following production grammar:



5. Conclusion: the evolution of auditory contrast

Evolution towards dispersion. The simulations show that a language with two or three sibilants evolves towards a *dispersed* system, i.e. one that has the 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.

Not necessarily encoded in the brain:

- No knowledge of auditory distances has to be encoded in the brain: we have shown that the diachronic emergence of sibilants in English and Polish can be explained without invoking dispersion constraints that are specific to inventory optimization.
- There are no markedness constraints (contra Prince & Smolensky). Markedness effects result from cue, articulatory, and sensorimotor constraints. These *can* be phonologized as language-specific structural constraints (NoCoda, AllFeetLeft).
- Richness of the Base is not in the underlying form (contra Prince & Smolensky), i.e. not in the brain. It is in the auditory form, i.e. in the outside world. It is the listener who filters the rich auditory input into language-specific allowed structures, by making use of her cue constraints and her structural constraints.

Prediction: everything else being equal, we expect a stable language with two or three sibilants to have the same auditory inventory as English and Polish. In reality, of course, the inventory will be influenced by the rest of the phonological system of the language, because *tout se tient*.

No teleology. As a corollary, we have also shown that a non-teleological theory of sound change can account for cases of diachronic contrast enhancement, a possibility that was argued against by Blevins (2004: 293–294).

References

Blevins, Juliette (2004). *Evolutionary phonology*. Oxford: Oxford University Press.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences* (University of Amsterdam) **21**: 43–58.

Boersma, Paul (1998). *Functional phonology: formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. In Robert Kirchner, Wolf Wikeley, and Joe Pater (eds.): *Papers in Experimental and Theoretical Linguistics*. Vol. **6**. Edmonton: University of Alberta. 24–35.

Boersma, Paul (2005a). Some listener-oriented accounts of hache-aspiré in French. *ROA* **730**.

Boersma, Paul (2005b). Prototypicality judgments as inverted perception. *ROA* **742**.

Boersma, Paul, and Silke Hamann (2004). The violability of backness in retroflex consonants. *ROA* **713**.

Carlton, Terrance R. (1991). *Introduction to the phonological history of the Slavic languages*. Columbus: Slavica.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, Stanka Fitneva, and Jacques Mehler (1999). Epenthetic vowels in Japanese: a perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance* **25**: 1568–1578.

Escudero, Paola, and Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.

Flemming, Edward (1995). *Auditory representations in phonology*. Doctoral dissertation, UCLA. [Published 2002 by Routledge, New York & London]

Forrest, Karen, Gary Weismer, Paul Milenkovic, and Ronald Dougall (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* **84**: 115–123.

Gordon, Matthew, Paul Barthmaier, and Kathy Sands (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* **32**: 141–174.

Jongman, Allard, Ratree Wayland, and Serena Wong (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* **108**: 1252–1263.

Kirchner, Robert (1998). *An effort-based approach to consonant lenition*. Doctoral dissertation, UCLA. [Published 2001 by Routledge, New York & London]

Liljencrants, Johan, and Björn Lindblom (1972). Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language* **48**: 839–862.

Lindblom, Björn (1986). Phonetic universals in vowel systems. In John Ohala and Jeri Jaeger (eds.) *Experimental phonology*. Orlando: Academic Press. 13–44.

McCarthy, John, and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. University of Massachusetts Occasional Papers **18**. Amherst, Mass.: Graduate Linguistic Student Association.

McQueen, James, and Anne Cutler (1997). Cognitive processes in speech perception. In William Hardcastle and John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell. 566–585.

Nowak, Paweł (to appear). The role of vowel transitions and frication noise in the perception of Polish sibilants. *Journal of Phonetics*.

Padgett, Jaye, and Marzena Zygis (2003). The evolution of sibilants in Polish and Russian. *ZAS Working Papers in Linguistics* **32**: 155–174.

Polivanov, Evgenij Dmitrievič (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* **4**: 79–96. [English translation: 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, and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. [Published 2004 by Blackwell, London]

Rospond, Stanisław (1971) *Gramatyka historyczna języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.

Sanders, Nathan (2003). *Opacity and sound change in the Polish lexicon*. Doctoral dissertation, University of California, Santa Cruz.

Steriade, Donca (1995). Positional neutralization. Unfinished manuscript, UCLA.

Stieber, Zdzisław (1952). *Rozwój fonologiczny języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe. [Translated 1968 into English by E. Schwartz as *The phonological development of Polish*. Michigan Slavic Materials **8**. Ann Arbor: University of Michigan]

Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes, and Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*. Somerville, Mass.: Cascadilla. 615–626.

Tesar, Bruce, and Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Zygis, Marzena, and Silke Hamann (2003). Perceptual and acoustic cues of Polish coronal fricatives. *Proceedings of the 15th International Conference of Phonetic Sciences*, Barcelona. 395–398.