Prototypicality judgments as inverted perception

Paul Boersma, May 17, 2005

In recent work (Boersma & Hayes 2001), Stochastic Optimality Theory has been used to model grammaticality judgments in exactly the same way as corpus frequencies are modelled, namely as the result of noisy evaluation of constraints ranked along a continuous scale. It has been observed, however, that grammaticality judgments do not necessarily reflect relative corpus frequencies: it is possible that structure A is judged as more grammatical than a competing structure B, whereas at the same time structure B occurs more often in actual language data than structure A. The present paper addresses one of these observations, namely the finding that 'ideal' forms found in experiments on *prototypicality judgments* often turn out to be peripheral within the corpus distribution of their grammatical category (Johnson, Flemming & Wright 1993). At first sight one must expect that Stochastic Optimality Theory will have trouble handling such observed discrepancies. The present paper, however, shows that a bidirectional model of phonetic perception and production (Boersma 2005) solves the paradox. In that model, corpus frequency reflects the *production process*, whereas prototypicality judgments naturally derive from a simpler process, namely the *inverted perception process*.

1 The /i/ prototype effect: prototypes are peripheral

A notorious example of the difference between grammaticality judgments and corpus frequencies is the "/i/ prototype effect" in phonology: if the experimenter asks a subject to choose the most /i/-like vowel from among a set of tokens that vary in their spectral properties, the subject tends to choose a very peripheral token, i.e. one with a very low first formant (e.g. 250 Hz) and a very high second formant (Johnson, Flemming & Wright 1993; Frieda, Walley, Flege & Sloane 2000). In actual speech, less extreme formant values (e.g. and F1 of 300 Hz) are much more common. Apparently, then, the token that the subject prefers is much more /i/-like than the average realization of the vowel /i/ is.

1.1 Why the /i/ prototype effect is a problem for linguistic models

The /i/ prototype effect has consequences for models of phonological grammar. The commonly assumed three-level grammar model, for instance, has trouble accounting for it. In this model, the *phonology* module maps an abstract underlying form (UF), for instance the lexical vowel |i|, to an equally discrete abstract surface form (SF), for instance /i/, and the *phonetic implementation* module subsequently maps this phonological SF to a continuous overt phonetic form (OF), which has auditory correlates, such as a value of the first formant, and articulatory correlates, such as a certain tongue height and shape. Such a grammar model can thus be abbreviated as UF→SF→OF.

The experimental prototypicality judgment task described above involves a mapping from the phonological surface form /i/ to an overt auditory first formant value, i.e. an SF→OF mapping. In the three-level grammar model, therefore, the natural way to account for this task is to assume that it shares the SF→OF mapping with the phonetic implementation process. If so, corpus frequencies (which result from phonetic implementation) should be the same as grammaticality judgments (whose best result is the prototype). Given that the two turn out to be different, Johnson, Flemming & Wright (1993) found the UF→SF→OF model wanting and proposed the model UF→SF→HyperOF→OF, where the additional intermediate representation HyperOF is a 'hyperarticulated' *phonetic target*. The prototypicality task, then, was proposed to tap HyperOF, whereas corpus frequencies reflect OF. The present paper shows, however, that if one distinguishes between articulatory and auditory representations at OF, the two tasks (production and prototypicality) involve different mappings, and the /i/ prototypicality effect arises automatically without invoking the additional machinery of an extra intermediate representation and an extra processing stratum.

2 A bidirectional constraint-based explanation of the /i/ prototype effect

This chapter presents a simple constraint-based model of the phonological grammar and of five phonological processes that are defined on this grammar. The account leads to an informal explanation for the /i/ prototype effect.

2.1 A grammar model with two phonological and two phonetic representations

The grammar model presented in Figure 1 is the Optimality-Theoretic model of "phonology and phonetic in parallel" (Boersma 2005).

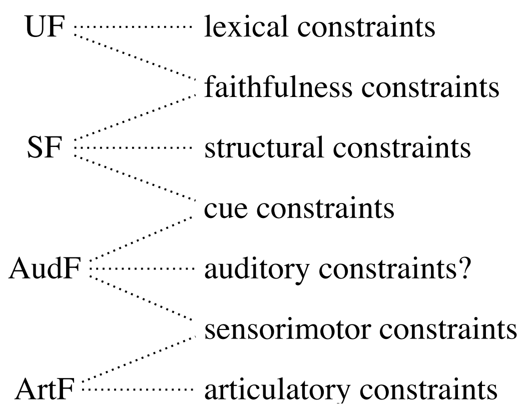

Fig. 1. The grammar model underlying bidirectional phonology and phonetics.

Figure 1 shows the four relevant representations and their connections. There are two separate phonetic forms: the *auditory form* (AudF) appears because it is the input to comprehension, and the *articulatory form* (ArtF) appears because it is the output of production. ArtF occurs below AudF because 9-month-old children can perceive sounds that they have no idea how to produce (for an overview, see Juszyk 1997); at this age, therefore, there has to be a connection from AudF to SF (or even to UF, once the lexicon starts to be built up) that cannot pass through ArtF; Figure 1 generalizes this for speakers of any age.

2.2 Linguistic processes are defined on the grammar

Figure 1 is not a processing model. Rather, linguistic and paralinguistic tasks have to be defined as processes that travel the representations in Figure 1 and are evaluated by the constraints that are visited on the way. Normal language use consists of two *linguistic* tasks: that of the listener (*comprehension*) and that of the speaker (*production*). This section describes the implementation of these two linguistic tasks; three *paralinguistic* tasks (including the prototypicality task), which can be regarded as simplified versions of the linguistic tasks, are described in the next section.

Boersma (2005) proposes that in the model of Figure 1 the linguistic task of comprehension is implemented as two consecutive mappings (cf. McQueen & Cutler 1997), as shown on the left in Figure 2.

The first mapping in comprehension is *perception* (also called *prelexical perception* or *phonetic parsing*). In general, perception is the process that maps continuous sensory information onto a more abstract mental representation. In phonology, perception is the process that maps continous auditory (and sometimes visual) information, i.e. AudF, onto a discrete phonological surface representation, i.e. SF. The shortest route from AudF to SF in Figure 1 determines what constraints evaluate this mapping: the relation between input (AudF) and output (SF) is evaluated with cue constraints (Escudero & Boersma 2003, 2004), and the output (SF) is evaluated with the familiar structural constraints known since the earliest Optimality Theory (OT; Prince & Smolensky 1993). This AudF→SF mapping is language-specific, and several aspects of it have been modelled in OT: categorizing auditory features to phonemes and autosegments (Boersma 1997, 1998 et seq.; Escudero & Boersma 2003, 2004), and building metrical foot structure (Tesar 1997, Tesar & Smolensky 2000, Apoussidou & Boersma 2004).

The second mapping in comprehension, shown at the top left in Figure 2, is that from SF to UF and can be called *recognition*, *word recognition*, or *lexical access*. In this mapping, the relation between input (SF) and output (UF) is evaluated with faithfulness constraints such as those familiar from two-level OT (McCarthy & Prince 1995), and the output (UF) is evaluated with lexical access constraints (Boersma 2001).

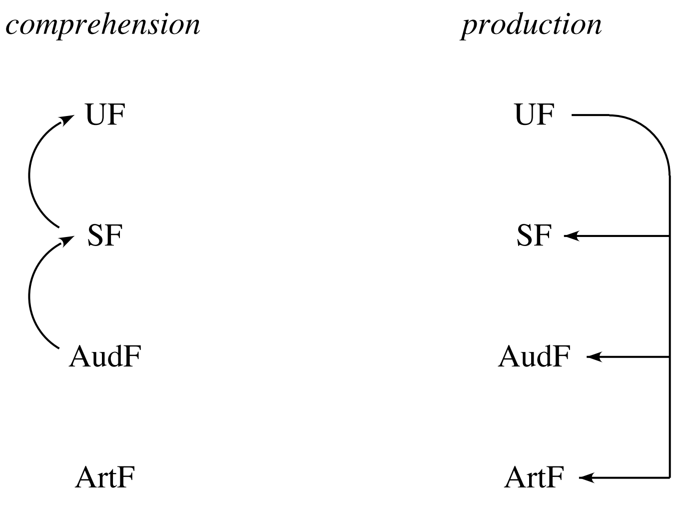

Fig. 2. The linguistic task of the listener, and that of the speaker.

Boersma (2005) proposes that in contradistinction to comprehension, production (shown at the right in Figure 2) consists of one single mapping from UF to ArtF, without stopping at any intermediate form as is done in comprehension. In travelling from UF to ArtF, the two representations SF and AudF are necessarily visited, so the production process must evaluates triplets of { SF, AudF, ArtF } in parallel. As can be seen from Figure 1, the evaluation of these triplets must be done with faithfulness constraints, structural constraints, cue constraints, sensorimotor constraints (which express the speaker's knowledge of how to pronounce a target auditory form, and of what a given articulation will sound like), and articulatory constraints (which express minimization of effort; see Boersma 1998, Kirchner 1998). According to Boersma (2005), the point of regarding phonological and phonetic production as parallel processes is that this can explain how discrete phonological decisions at SF can be influenced by gradient phonetic considerations such as salient auditory cues at AudF (e.g. Steriade 1995) and articulatory effort at ArtF (e.g. Kirchner 1998).

2.3 Experimental tasks are paralinguistic processes

Experimental tasks in the phonetics laboratory are often designed to reflect only a part of one of the linguistic processes shown in Figure 2. The present section addresses the three tasks that are relevant for explaining the /i/ prototype effect, namely the *phoneme categorization task*, the *phoneme production task*, and the *phoneme prototypicality task*.

In the experimental task of phoneme categorization the participant is asked to classify a given stimulus as one of the phonemes of her language, for instance to classify a synthetic vowel with a known F1 of 360 Hz as either the vowel /ɪ/ or the vowel /i/. Such an experiment tends to be set up in such a way that the influence of the lexicon is minimized, for instance by presenting the response categories as semantically empty vowel symbols (e.g. "a", "e", "i", "o", "u" for Spanish listeners) or as equally accessible lexical items (e.g. "ship" and "sheep" for English listeners).[1] In the former case, UF may not be accessed at all; in the latter case, the SF→UF mapping may be equally easy for both categories; in both cases, the influence of the lexicon may be ignored, so that the task can be abbreviated as in Figure 3 (left). The only constraints that are relevant for this mapping are the cue constraints and the structural constraints.

In the experimental task of phoneme production the participant is asked to pronounce either a nonsense word or a word with 'no' phonology, i.e. where SF is identical to UF (no faithfulness violations), such as English *hid* or *heed*. In both cases, the influence of the lexicon can again be ignored, so the task can be abbreviated as in Figure 3 (right). The relevant constraints will be the cue constraints, the sensorimotor constraints, and the articulatory constraints.

In the experimental task of prototypicality judgments the participant is given an SF, as in the phoneme production task, and asked to choose an AudF, similar to those in the phoneme categorization task. Since this task involves neither the lexicon nor any actual articulation, it can be abbreviated as Figure 3 (middle). The only relevant constraints are the cue constraints (if auditory constraints do not exist).

[1] Unless the very point of the experiment is to investigate the influence of the lexicon on prelexical perception. By the way, if such influences turn out to exist (e.g. Ganong 1980, Samuel 1981), the comprehension model in Figure 2 will have to be modified in such a way that perception and recognition work in parallel. In that case, however, the phoneme categorization task will still look like that in Figure 3, and the results of the present paper will still be valid.

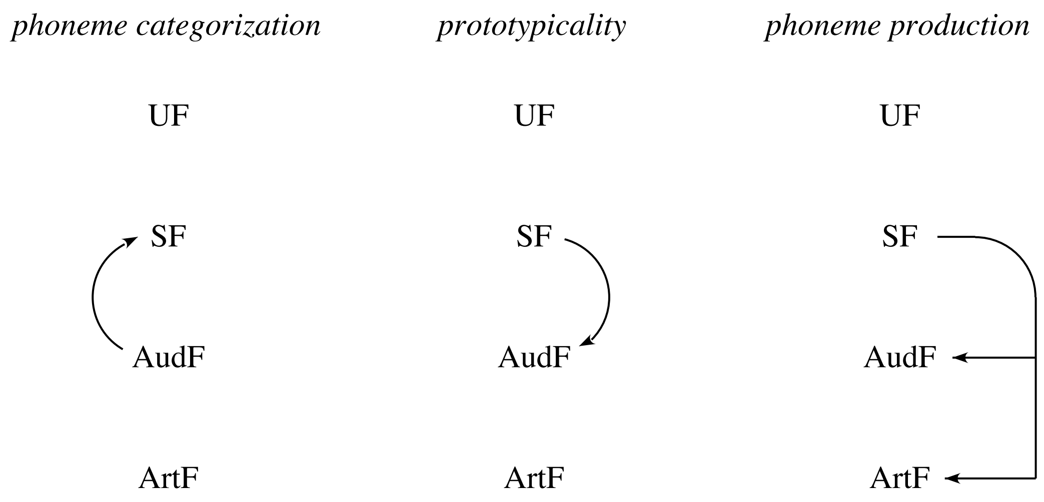

Fig. 3. Three laboratory tasks.

2.4 The informal explanation

The fact that the prototypicality task yields a different result than the phoneme production task can be attributed to the difference between the relevant two processes in Figure 3: in the production task, constraints on articulatory effort do play a role, in the prototypicality task they do not. This is a robust effect that seems to withstand conscious manipulation: even if listeners are asked to choose the auditory form that they would say themselves, they respond with the prototype, not with the form they would really produce (Johnson, Flemming & Wright 1993).

The result of the involvement of articulatory constraints in the phoneme production task is that peripheral tokens of /i/ may be outruled because they are too effortful, e.g. because they require too much articulatory precision, whereas tokens closer to the easiest vowel articulation, perhaps [ə], do not violate any high-ranked articulatory constraints.

3 A formalization in Optimality Theory

While the explanation presented informally in §2.4 would work for any constraint-based theory of bidirectional phonology and phonetics in parallel, this chapter formally shows that it works for the particular constraint-based framework of Stochastic Optimality Theory. The point of this exercise is not only to provide a rigorous illustrative example, so as to achieve descriptive adequacy, but also to propose an explanation of the acquisition of the relevant part of the grammar in terms of an initial state and a learning path, so as to achieve explanatory adequacy and to show that the resulting grammar is stable.

3.1 Formalizing phoneme categorization and its acquisition

As seen in Figure 3 (left), phoneme categorization can be seen as involving prelexical perception only, i.e. as a mapping from an auditory form to a phonological surface form. For the case of the /i/ prototype effect, it is relevant to look at auditory events that are likely to be perceived as /i/ or as one of its neighbours in the vowel space, such as /ɪ/ or /e/. Thus, the auditory form (AudF) is a combination of an F1 and an F2 value, and the surface form (SF) is a vowel segment such as /i/ or /e/. This section shows how the AudF→SF mapping is handled with an Optimality-Theoretic grammar that contains cue constraints, which evaluate the relation between AudF and SF, and structural constraints, which evaluate the output representation SF.

For simplicity I discuss the example of a language with three vowels, /a/, /e/ and /i/, in which the only auditory distinction between these vowels lies in their F1 values. Suppose that the speakers realize these three vowels most often with F1 values of 700 Hz, 500 Hz, and 300 Hz, respectively, but that they also vary in their realizations. If this variation can be modelled with Gaussian curves with standard deviations of 60 Hz, the distributions of the speakers' productions will look as in Figure 4.

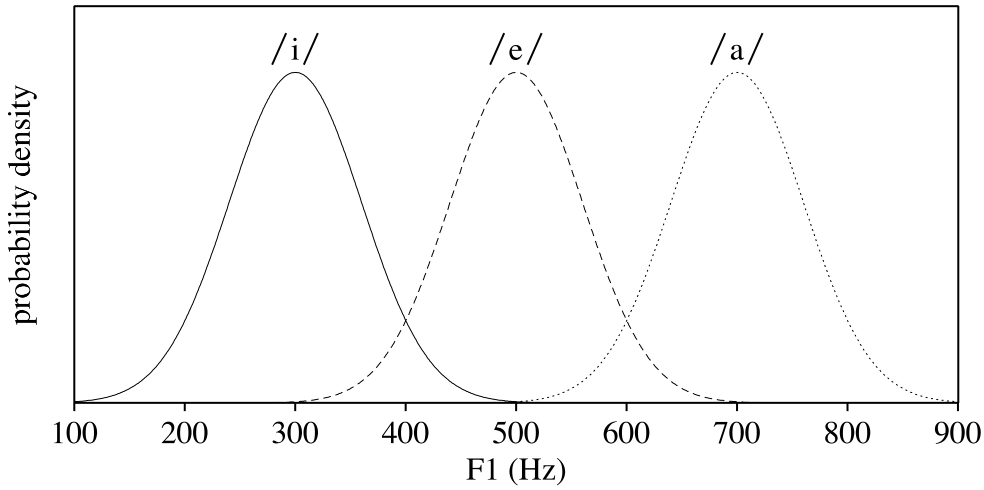

Fig. 4. Production distributions of three vowels.

Now how do listeners classify incoming F1 values, i.e. to which of the three categories /a/, /e/ and /i/ do they map a certain incoming F1 value *x*? This mapping can be handled by a family of negatively formulated Optimality-Theoretic cue constraints, which can be expressed as "if the auditory form contains an F1 of *x* Hz, the corresponding vowel in the surface form should not be *y*" (Escudero & Boersma 2003, 2004).[2] These cue constraints exist for all F1 values between 100 and 900 Hz and for all three vowels. Examples are given in (1).

(1) *Cue constraints for mapping F1 values to vowel categories*

"an F1 of 340 Hz is not /a/"
"an F1 of 340 Hz is not /e/"
"an F1 of 340 Hz is not /i/"
"an F1 of 539 Hz is not /i/"

[2] There are two reasons for the negative formulation of these constraints. First, a positive formulation would simply not work in the case of the integration of multiple auditory cues (Boersma & Escudero 2004). Second, the negative formulation allows these constraints to be used in both directions (comprehension and production), as is most clearly shown by the fact that they can be formulated symmetrically as *[*x*]/*y*/ (Boersma 2005). The former reason is not relevant for the present paper, but the second is, because the same cue constraints are used in the next two sections.

The second type of constraints involved in the AudF→SF mapping are the structural constraints that evaluate the output SF. In the present case, they could be something like */a/, */e/ and */i/. I assume that all three vowels are equally perfectly licit phonemes of the language, so that these constraints must be ranked low. I ignore them in the rest of this paper, so that phoneme categorization is handled solely by the cue constraints.[3]

The ranking of the cue constraints results from lexicon-driven perceptual learning (Boersma 1997, 1998; Escudero & Boersma 2003, 2004): the learner hears an auditory event drawn from the environmental distributions in Figure 4 and classifies it as a certain vowel, and the lexicon subsequently tells her which vowel category she should have perceived. This type of learning assumes that the acquisition process contains a period in which the listener already knows that the language has three vowel categories and in which all her lexical representations are already correct. If such a learner misclassifies a speaker's intended /pit/ as /pet/, her lexicon, which contains the underlying form |pit|, will tell her that she should have perceived /pit/ instead. When detecting an error in this way, the learner will take action by changing the ranking of some constraints. Suppose that at some point during acquisition some of the constraints are ranked as in (2). The learner will then perceive an incoming F1 of 380 Hz as the vowel /e/, as indicated by the pointing finger in (2). We can also read from (2) that 320 Hz will be perceived as /i/, and 460 Hz as /e/.

(2) *Learning to perceive vowel height*

[380 Hz] (UF = |i|)	320 Hz not /a/	380 Hz not /a/	460 Hz not /i/	320 Hz not /e/	460 Hz not /a/	380 Hz not /i/	380 Hz not /e/	320 Hz not /i/	460 Hz not /e/
/a/		*!							
☞ /e/							←*		
√ /i/						*!→			

If the lexicon now tells the learner that she should have perceived /i/ instead of /e/, she will regard this as the correct adult SF, as indicated by the check mark in (2). According to the Gradual Learning Algorithm for Stochastic OT (Boersma 1997, Boersma & Hayes 2001), the learner will take action by raising the ranking value of all the constraints that prefer the adult form /i/ to her own form /e/ (here only "380 Hz is not /e/") and by lowering the ranking value of all the constraints that prefer /e/ to /i/ (here only "380 Hz is not /i/"). These rerankings are indicated by the arrows in (2).

To see what kind of final perception behaviour this procedure leads to, I ran a computer simulation analogous to the one by Boersma (1997). A virtual learner has 243 constraints (F1 values from 100 to 900 Hz in steps of 10 Hz, for all three vowel categories), all with the same initial ranking value of 100.0. The learner then hears 10 million F1 values randomly drawn from the distributions in Figure 4, with an equal probability of 1/3 for each vowel. She is subjected to the learning procedure exemplified in (2), with full knowledge of the lexical form, with an evaluation noise of 2.0, and with a *plasticity* (the

[3] Auditory constraints, if they exist, evaluate the input and cannot therefore distinguish between the candidates.

amount by which ranking values rise or fall when a learning step is taken) of 0.01. The result is shown in Figure 5.

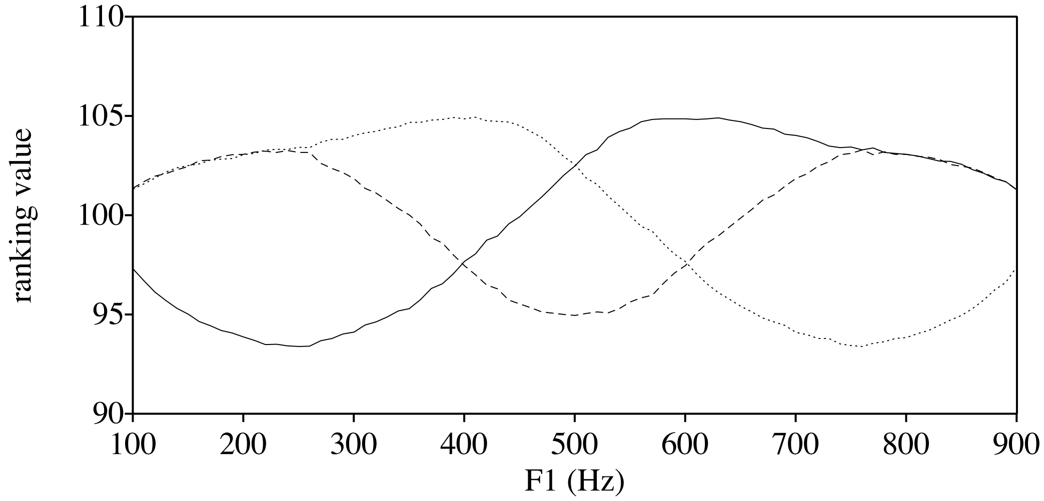

Fig. 5. The final ranking of "an F1 of *x* Hz is not /*vowel*/", for the vowels /i/ (solid curve), /e/ (dashed curve), and /a/ (dotted curve).

The figure is to be read as follows. F1 values below 400 Hz will mostly be perceived as /i/, since in that region the constraint "an F1 of *x* Hz is not /i/" (the solid curve) is ranked lower than the constraints "an F1 of *x* Hz is not /e/" (the dashed curve) and "an F1 of *x* Hz is not /a/" (the dotted curve). Likewise, F1 values above 600 Hz will mostly be perceived as /a/, and values between 400 and 600 Hz mostly as /e/. For every F1 value the figure shows us not only the most often perceived category but also the degree of variation. Around 400 Hz, /i/ and /e/ perceptions are equally likely. Below 400 Hz it becomes more likely that the listener will perceive /i/, and increasingly so when the distance between the curves for /i/ and /e/ increases. This distance is largest for F1 values around 250 Hz, where there are 99.8% /i/ perceptions and only 0.1% perceptions of /e/ and /a/ each. Below 250 Hz, the curves approach each other again, leading to more variation in categorization.

A detailed explanation of the shapes of the curves in terms of properties of the Gradual Learning Algorithm (approximate probability matching between 250 and 750 Hz, and low corpus frequencies around 100 and 900 Hz) is provided at the end of the next section, where the shapes of the curves are related to the behaviour of prototypicality judges.

3.2 Formalizing the prototypicality task

As seen in Figure 3 (middle), the prototypicality task can be seen as a mapping from a phonological surface form to an auditory form, without the involvement of an articulatory form. This section shows how this SF→AudF mapping is handled with the same Optimality-Theoretic cue constraints as phoneme categorization. From Figure 1, we can see that auditory constraints might be involved in evaluating the output of this mapping, but given that we do not know whether such constraints (against loud and unpleasant noises?) are relevant at all for phonology, I ignore them here.

The mapping from SF to AudF in the prototypicality task is thus entirely handled by the cue constraints. For the listener simulated in the previous section, these constraints are ranked as in Figure 5. The ranking of the constraints for /i/ has been copied from the solid curve in Figure 5 to the top row in tableau (3). In Figure 5, for instance, the bottom of the /i/ curve lies at an F1 of 250 Hz. In (3) this is reflected by the bottom ranking of "250 Hz is not /i/". In (3) we also see that as the F1 goes up or down from 250 Hz, the constraint against perceiving this F1 as /i/ becomes higher ranked, just as in Figure 5.

(3) *The auditory F1 value that gives the best /i/*

/i/	320 Hz not /i/	310 Hz not /i/	170 Hz not /i/	180 Hz not /i/	300 Hz not /i/	190 Hz not /i/	290 Hz not /i/	200 Hz not /i/	280 Hz not /i/	210 Hz not /i/	270 Hz not /i/	230 Hz not /i/	220 Hz not /i/	240 Hz not /i/	260 Hz not /i/	250 Hz not /i/
[170 Hz]			*!													
[180 Hz]				*!												
[190 Hz]						*!										
[200 Hz]								*!								
[210 Hz]										*!						
[220 Hz]													*!			
[230 Hz]												*!				
[240 Hz]														*!		
☞ [250 Hz]																*
[260 Hz]															*!	
[270 Hz]											*!					
[280 Hz]									*!							
[290 Hz]							*!									
[300 Hz]					*!											
[310 Hz]		*!														
[320 Hz]	*!															

With the ranking shown in Figure 5 and tableau (3), and with zero evaluation noise, the listener will choose an F1 of 250 Hz as the optimal value for /i/. This is more peripheral (more towards the edge of the F1 continuum) than the most often heard /i/, which has an F1 of 300 Hz according to Figure 4. The size of the effect (50 Hz) is comparable to the effect found by Johnson, Flemming & Wright (1993) and Frieda, Walley, Flege & Sloane (2000). Of course, this simulated value of 50 Hz depends on several assumptions, such as an initial equal ranking for all the constraints, which is probably unrealistic (for a more realistic proposal based on a period of distributional learning before lexicon-driven learning, see Boersma, Escudero & Hayes 2003). The parameter that determines the size of the effect in the present simulation is the standard deviation of the F1 values in the environmental distribution in Figure 1, which was 60 Hz. With different standard deviations, a different effect size is expected.

The result of 250 Hz in tableau (3) was based on a categorical ranking of the constraints. In the presence of evaluation noise the outcome will vary. If the evaluation noise is 2.0, i.e. the same as during the learning procedure of the previous section, the outcome for the listener of Figure 5 and tableau (3) will vary as in Figure 6, which was assembled by computing the outcomes of 100,000 tableaus like (3).

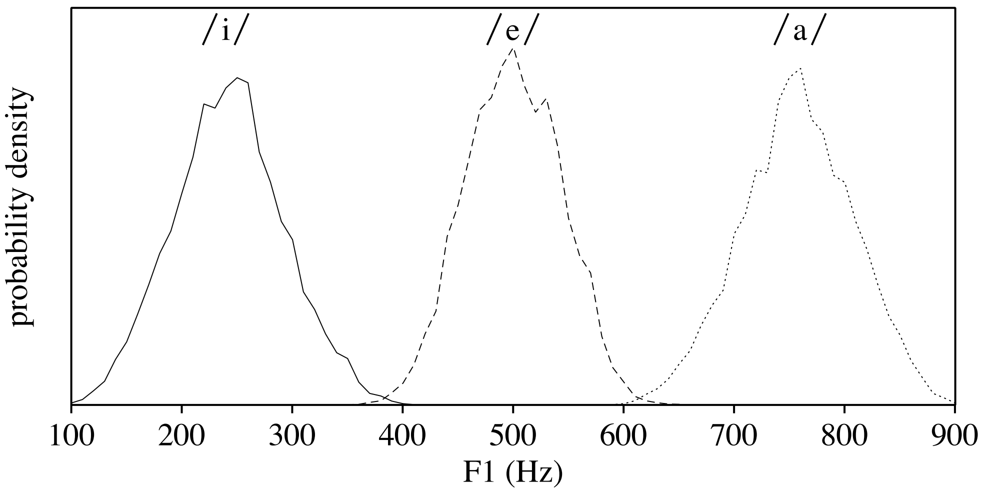

Fig. 6. Prototypicality distributions for the three vowels.

The differences between environmental F1 values and prototypicality judgments seen when comparing figures 4 and 6 are very similar to the production/perception differences in the experiments by Johnson, Flemming & Wright (1993) and Frieda, Walley, Flege & Sloane (2000).

The conclusion is that if the prototypicality task uses the same constraint ranking as phoneme categorization, auditorily peripheral segments will be judged best if their auditory values are extreme, because cue constraints have automatically been ranked lower for extreme auditory values than for more central auditory values. The question that remains is: how has the /i/ curve in Figure 5 become lower at 250 Hz than at 300 Hz? The answer given by Boersma (1997) is the probability-matching property of the Gradual Learning Algorithm: the ultimate vertical distance between the /i/ and /e/ curves for a given F1 is determined (after learning from a sufficient amount of data) by the probability that that F1 reflects an intended /i/ rather than an intended /e/; given that an F1 of 250 Hz has a smaller probability of having been intended as /e/ than an F1 of 300 Hz, the vertical distance between the /i/ and /e/ curves grows to be larger at 250 Hz than at 300 Hz, providing that the learner is given sufficient data. With the Gradual Learning Algorithm and enough input, the prototypicality judge will automatically come to choose the F1 token that is least likely to be perceived as anything else than /i/.[4]

[4] This goal of choosing the least confusing token was proposed by Lacerda (1997) as the driving force behind the prototypicality judgment. He did not propose an underlying mechanism, though. See also §4.

There are two reasons why the prototype does not have an even lower F1 than 250 Hz. The first reason, which can be illustrated with the simulation, is that there are simply not enough F1 values of, say, 200 Hz to allow the learner to reach the final state of a wide separation between the /i/ and /e/ curves; for the simulated learner, Figure 5 shows that even 10 million inputs did not suffice. The second reason, not illustrated by the simulation, is that in reality the F1 values are not unbounded. Very low F1 values are likely to be perceived as an approximant, fricative, or stop rather than as /i/. Even within the vowel space, this effect can be seen at the other end of the continuum: one would think that the best token for /a/ would have an extremely high F1, but in reality an F1 of, say, 3000 Hz will be perceived as /i/, because the listener will reinterpret it as an F2 with a missing F1.

3.3 Formalizing phoneme production

Now that we have seen how inverted perception accounts for the /i/ prototype effect, we still have to see how it is possible that the same peripheral values are not used in the phoneme production task. Presumably, after all, the learner as a speaker will grow to match the modal F1 value of 300 Hz that she finds in her environment (if sound change can be ignored).

The answer is shown in Figure 3 (right): the phoneme production task takes an SF as its input (as does the prototypicality task), but has to generate both an auditory form and an articulatory form as its output. Similarly to the protypicality task, the production process will have to take into account the cue constraints, but unlike the prototypicality task, the production process will also have to take into account sensorimotor constraints and articulatory constraints.

Tableau (4) shows how the phonological surface form /i/ is produced phonetically. The cue constraints are still ranked exactly as in (3). Every candidate cell, however, now contains a pair of phonetic representations: articulatory and auditory. The articulatory part of each candidate shows the gestures needed for articulating [i]-like sounds. For simplicity I assume that the main issue is the precision with which the tongue has to be bulged towards the palate, and that more precision yields lower F1 values, e.g. a precision of ‘26’ yields an F1 of 240 Hz whereas a precision of ‘17’ yields an F1 of 330 Hz. These precison values are evaluated by articulatory constraints that are ranked by the amount of effort involved, i.e. the constraints “the precision should not be greater than 26” has to outrank the constraint “the precision should not be greater than 17”.

The sensorimotor constraints are missing from tableau (4). This is because for purposes of simplicity I assume here that the relation between articulatory and auditory form is fixed, i.e., the speaker has a fully proficient view of what any articulation will sound like and of how any auditory event can be implemented articulatorily. The candidate $[240\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=26]_{\text{Art}}$, for instance, occurs in (4) because it only violates the low-ranked sensorimotor constraint $*[240\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=26]_{\text{Art}}$, whereas candidates like $[240\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=22]_{\text{Art}}$ and $[270\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=26]_{\text{Art}}$ violate the high-ranked sensorimotor constraints $*[240\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=22]_{\text{Art}}$ and $*[270\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=26]_{\text{Art}}$ and are therefore ignored in tableau (4). By making this simplification, we can regard the relationship between AudF and ArtF as fixed, so that only the cue constraints and the articulatory constraints determine the speaker’s behaviour.

(4) *Producing vowel height*

/i/	prec not > 26	320 Hz not /i/	prec not > 24	310 Hz not /i/	prec not > 22	170 Hz not /i/	prec not > 20	180 Hz not /i/	prec not > 18	300 Hz not /i/	prec not > 16	190 Hz not /i/	290 Hz not /i/	200 Hz not /i/	280 Hz not /i/	210 Hz not /i/	270 Hz not /i/	230 Hz not /i/	220 Hz not /i/	240 Hz not /i/	260 Hz not /i/	250 Hz not /i/
$[170\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=33]_{\text{Art}}$	*!		*		*	*	*		*		*											
$[180\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=32]_{\text{Art}}$	*!		*		*		*	*	*		*											
$[190\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=31]_{\text{Art}}$	*!		*		*		*		*		*	*										
$[200\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=30]_{\text{Art}}$	*!		*		*		*		*		*			*								
$[210\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=29]_{\text{Art}}$	*!		*		*		*		*		*					*						
$[220\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=28]_{\text{Art}}$	*!		*		*		*		*		*								*			
$[230\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=27]_{\text{Art}}$	*!		*		*		*		*		*							*				
$[240\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=26]_{\text{Art}}$			*!		*		*		*		*									*		
$[250\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=25]_{\text{Art}}$			*!		*		*		*		*											*
$[260\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=24]_{\text{Art}}$					*!		*		*		*										*	
$[270\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=23]_{\text{Art}}$					*!		*		*		*						*					
$[280\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=22]_{\text{Art}}$							*!		*		*				*							
$[290\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=21]_{\text{Art}}$							*!		*		*		*									
☞ $[300\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=20]_{\text{Art}}$									*	*	*											
$[310\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=19]_{\text{Art}}$				*!					*		*											
$[320\ \text{Hz}]_{\text{Aud}}$ $[\text{prec}=18]_{\text{Art}}$		*!									*											

The result of the ranking in (4) is that the auditory-articulatory pair [F1 = 300 Hz]$_{Aud}$ [prec=20]$_{Art}$ wins. Forms with a lower F1 are too effortful, whereas forms with a higher F1 are too confusable.

The result of the phoneme production task is thus very different from that of the prototypicality task. The difference between the two tasks can be reduced to the presence of the articulatory constraints in the production task and their absence in the prototypicality task.

3.4 The formal explanation for the /i/ prototype effect

With the tableaus and simulation in §3.1 through §3.3 the /i/ prototype effect can now be explained in more formal terms than in §2.4. The simulation in §3.1 explains the fact that the F1 in the prototypicality task was 50 Hz lower than the modal F1 in the learner's environment, while the difference between the tableaus in §3.2 and §3.3 explains the fact that the F1 in the production task was 50 Hz higher than in the prototypicality task. One can say that the *prototypicality effect* is –50 Hz and the *articulatory effect* is +50 Hz, and that the two effects cancel out.

The fact that the two effects cancel out in the example of §3.1 to §3.3 is due to my arbitrary choices for the ranking of the articulatory constraints in tableau (4). Had I ranked these constraints higher, the candidate [310 Hz] might have won in tableau (4); had I ranked them lower, the candidate [290 Hz] might have won. Either alternative would have led to a predicted shift in F1 from one generation of speakers to the next. The actual ranking in (4) was chosen in such a way that the two effects cancel out exactly, so that the production distributions stay stable over the generations.

To sum up: three F1 values for /i/ have been addressed: the modal F1 of the first generation, the prototypical F1 for the second generation, and the modal F1 of the second generation. If the prototypicality and articulatory effects cancel out (as they do in reality, if there is no sound change), the first and third of these F1 values will be identical, and the prototype F1 will be the odd one out and the most conspicuous to researchers. Its difference from the two modal F1 values has been accounted for in §3.2 and §3.3, respectively.

3.5 Stability over generations: a coincidence?

The really surprising observation is now no longer the fact that the prototypicality task leads to a different F1 than the modal F1 produced by the first and second generations, but the fact that the modal F1 of the second generation is identical to that of the first.

The question of the stability of the production distribution could be answered with the help of Kirchner's (1998:288) proposal that the ranking of articulatory constraints is fixed, namely simply a function of articulatory effort alone. Imagine that this fixed ranking is the one in tableau (4), but that a learner is confronted with a environmental distribution with a modal F1 of 280 instead of 300 Hz. The F1 of the prototype /i/ will shift down, but not by 20 Hz, because the influence of /e/ tokens decreases; it may shift from 250 to, say, 235 Hz. The modal produced F1 will also shift, but not by 15 Hz, because the articulatory constraints do not shift; it may shift to, say, 290 Hz. Within one generation, therefore, the modal F1 for /i/ will rise from 280 to 290 Hz, and in a couple of generations more it will be very close to 300 Hz. An analogous thing will happen if the environmental distribution has a modal F1 of 320 Hz: it will move to 300 Hz in three generations or so. Given a fixed ranking of the articulatory constraints, therefore, every language will reach the same equilibirum: 300 Hz is the only stable F1 value possible, as long as everything else remains equal to the case of Figure 4. Other possible explanations for cross-generational stability involve the various learning algorithms for production in the parallel phonological-phonetic model (Boersma 2005), but these are far outside the scope of the present paper.

4 Comparison with earlier accounts

Tableaus (3) and (4) automatically predict that, if the child is given enough time to learn even the rare overt forms, the best auditory form is one that is less likely to be perceived as /e/ than the modal F1 value for /i/ is, and that articulatory constraints lead to a higher F1 value in production. This can be explained within grammar models in which ArtF can influence AudF, because in such models the resulting AudF will be different according to whether an ArtF has to be evaluated (as in the phoneme production task) or not (as in the prototypicality task). Such models include the one exemplified in the present paper, namely Boersma's (2005) parallel model of phonology and phonetics, where production is modelled as UF $\rightarrow$ { SF, AudF, ArtF }, but they also include Boersma's (1998) earlier listener-oriented grammar model, where production is modelled as UF $\rightarrow$ (ArtF$\rightarrow$AudF$\rightarrow$SF). They do *not* include forward modular models of production of the type UF $\rightarrow$ SF $\rightarrow$ AudF $\rightarrow$ ArtF, because in such serial models articulatory restrictions cannot influence the auditory form.

The prototypicality proposal by Johnson, Flemming & Wright (1993) is an example of a serial production model. Presumably, their model would abbreviate the prototypicality task as SF $\rightarrow$ HyperOF, and the phoneme production task as SF $\rightarrow$ HyperOF $\rightarrow$ OF. Since the presence vs. absence of the 'later' representation OF has no way of influencing the form of the 'earlier' representation HyperOF, this 'hyperarticulated phonetic target' has to contain a peripheral F1 value of 250 Hz that is independent from the experimental task. The authors provide no conclusive independent evidence for the existence of such a representation, whereas the representations AudF and ArtF proposed in the present paper are independently needed to serve as the input to comprehension and the output of production.

The prototypicality proposal by Frieda, Walley, Flege & Sloane (2000) invokes an extra representation as well, namely the 'prototype'. For the existence of this level of representation Kuhl (1991) gave some independent evidence, namely the 'perceptual magnet' effect. However, this effect can be explained without invoking prototypes. This has first been shown for lexically labelled exemplars by Lacerda (1995), and even in models of pure distributional learning without lexical labels, perceptual warping automatically emerges as an epiphenomenon, as has been shown for neural maps by Guenther & Gjaja (1996) and for Optimality Theory by Boersma, Escudero & Hayes (2003). With Occam's razor, explanations without these poorly supported prototypes have to be preferred.

The prototypicality proposal by Lacerda (1997) derives goodness judgments from the activations of categories in an exemplar model of phonology. However, the auditory token that generates the highest activation for /i/ is still the modal auditory form; the best prototype is only derived indirectly as the auditory form that has the highest activation for /i/ *relative to* its activation for other vowel categories. This proposal thus does choose the least confusable auditory form, but does not provide an automatic underlying mechanism such as the one that necessarily follows from the task in Figure 3 (middle).

Finally, the results derived in this paper could equally well have been derived by formalizing the grammar and task models in Figures 1 to 3 within a framework that does not rely on constraint ranking but on constraint weight addition, such as Harmony Grammar (Legendre, Miyata & Smolensky 1990ab).

5 Conclusion

The present paper has offered formal explanations for two facts, namely that a prototype (by being less confusable) is more peripheral than the modal auditory form in the listener's language environment, and that a prototype (by not being limited by articulatory restrictions) is more peripheral than the modal auditory form that the listener herself will produce. Given the representation-and-constraints model of Figure 1, the only assumptions that led to these formal explanations were that representations are evaluated only if they are necessarily activated (Figure 3) and that in production processes (Figure 2, right; Figure 3, right) the output representations are evaluated in parallel, so that 'later' representations can influence 'earlier' representations.

The explanations provided here for a phonetic example may well extend to other areas of linguistics. The place where grammaticality judgments have most often been investigated is that of syntactic theory. One can imagine that the corpus frequency of constructions that are informed by speaker-based requirements at Phonetic Form is greater than would be expected on the basis of grammaticality judgments in a laboratory reading task, which may only activate Logical Form. This, however, is worth a separate investigation.

References

Apoussidou, Diana, & Paul Boersma (2004). Comparing two Optimality-Theoretic learning algorithms for Latin stress. *WCCFL* **23**: 29–42.

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

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. Doctoral dissertation, University of Amsterdam.

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

Boersma, Paul (2005). Some listener-oriented accounts of hache aspiré in French. Ms., University of Amsterdam. [Rutgers Optimality Archive **730**, http://roa.rutgers.edu]

Boersma, Paul, & Paola Escudero (2004). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. *Rutgers Optimality Archive* **684**.

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. *Proceedings of the 15th International Congress of Phonetic Sciences*. 1013–1016.

Boersma, Paul, and Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* **32**: 45–86.

Escudero, Paola, and Paul Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser, and Alexander Williams (eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* **8.1**: 71–85.

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

Frieda, Elaina M., Amanda C. Walley, James E. Flege, and Michael E. Sloane (2000). Adults' perception and production of the English vowel /i/. *Journal of Speech, Language, and Hearing Research* **43**: 129–143.

Ganong, William F. III (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* **6**: 110–125.

Guenther, Frank H. & Marin N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *JASA* **100**: 1111-1121.

Johnson, Keith, Edward Flemming, and Richard Wright (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language* **69**: 505–528.

Jusczyk, Peter (1997). *The discovery of spoken language*. Cambridge, Mass.: MIT Press.

Kirchner, Robert (1998). *Lenition in phonetically-based Optimality Theory*. PhD dissertation, UCLA.

Kuhl, Patricia K. (1991). Human adults and human infants show a "perceptual magnetic effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* **50**: 93–107.

Lacerda, Francisco (1995). The perceptual magnet effect: an emergent consequence of exemplar-based phonetic memory. *Proceedings of the XIIIth International Congress of Phonetic Sciences* **2**: 140–147.

Lacerda, Francisco (1997). Distributed memory representations generate the perceptual-magnet effect. Ms., Institute of Linguistics, Stockholm University.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky (1990a). Harmonic Grammar — A formal multi-level connectionist theory of linguistic well-formedness: theoretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Sciences*. Cambridge, Mass.: Lawrence Erlbaum. 388–395.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky (1990b). Harmonic Grammar — A formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Sciences*. Cambridge, Mass.: Lawrence Erlbaum. 884–891.

McCarthy, John, & Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*. University of Massachusetts Occasional Papers **18**. Amherst, Mass.: Graduate Linguistic Student Association. 249–384.

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

Prince, Alan, & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Technical Report TR-**2**, Rutgers University Center for Cognitive Science.

Samuel, Arthur G. (1981). The role of bottom-up confirmation in the phonemic-restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance* **7**: 1124–1131.

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

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