

The Linguistic Perception of SIMILAR L2 sounds

Paola Escudero, University of Amsterdam

Abstract

In this article, I discuss a linguistic model for explaining second-language (L2) sound perception, which is a phenomenon that has commonly been modelled within the disciplines of phonetics and psycholinguistics. This linguistic model will be applied to the learning of SIMILAR L2 sounds. This L2 learning scenario refers to the acquisition of the knowledge involved in the perception of L2 sounds that are phonologically equivalent but yet phonetically different from the acoustically closest sounds in the learner's first language (L1). In the introduction, I argue, based on phonetic and psycholinguistic grounds, that speech perception is a language-specific phenomenon that should be brought into the domain of phonological modelling. Additionally, I propose a number of characteristics which incorporate phonological, phonetic, and psycholinguistic modelling and which should be found in a comprehensive and explanatory adequate model for sound perception. In § 2, I demonstrate that the Linguistic Perception (LP) model complies with these criteria. Crucially, I show that the L1 acquisition component of the LP model is shown to constitute a successful proposal for the mechanisms involved in learning to perceive L1 sounds. In § 3, I show how the L2 version of the LP model successfully describes, explains, and predicts the learning of SIMILAR L2 sounds. Specifically, the model predicts that listeners are optimal perceivers of their native language and that beginning L2 learners start with a copy of their L1 optimal perception. These two predictions are confirmed by the perception of /æ/ and /ɛ/ by monolingual Canadian English (CE) and Canadian French (CF) listeners and by the L2 perception of beginning CE learners of CF. Further, the model predicts that learners will adjust their initial L2 perception by means of the same mechanism used by L1 learners. This developmental prediction is confirmed by the gradual shifting of the category boundary between /æ/ and /ɛ/ in CE learners of CF. Finally, the L2LP model hypothesizes that both L1 and L2 can be optimal because they are handled by two separate grammars. The data demonstrate that CE learners of CF have differential perception systems for their L1 and L2. In sum, it is shown that this model provides the currently most comprehensive description, explanation, and prediction of L2 sound perception. It successfully incorporates an L2 phenomenon which was commonly regarded as phonetic or psycholinguistic within the domain of phonology, a modelling proposal which follows the tradition started by Escudero and Boersma (2004).

1 Introduction

In the following section, I discuss the Linguistic Perception (LP) model for general sound perception and L1 acquisition. In § 3, I illustrate the L2 Linguistic Perception (L2LP) model with the learning of L2 SIMILAR sounds, which are L2 sounds that are phonologically equivalent but yet phonetically different from the sounds in the learner's first language (L1) that are acoustically most similar. In this introductory section, I argue why we need a linguistic model for sound perception (§ 1.1) and discuss the characteristics that such a model should have to adequately describe and explain the phenomenon at hand (§ 1.2)

1.1 Why a phonological model for sound perception?

It is widely accepted that the human perceptual system organizes raw sensory input into abstract mental representations. For speech perception, this means that the listener converts raw auditory input into linguistic units such as vowels and consonants, as illustrated in Figure

1. For instance, English listeners will categorize a vowel with a short duration, a high first formant (F1), and a high second formant (F2) as the vowel /æ/ in the word *cat*, probably because English speakers tend to pronounce the vowel in *cat* with those same properties.

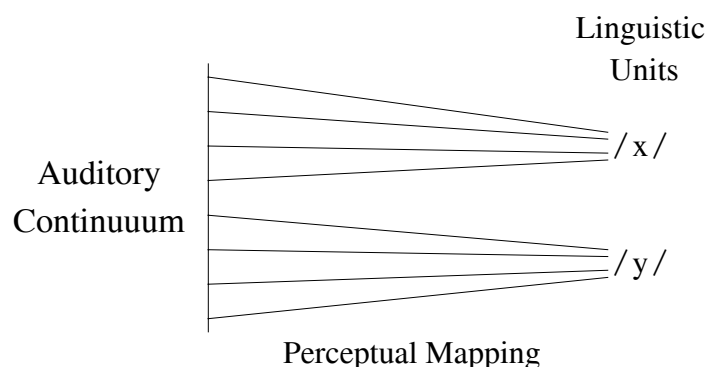


Fig. 1: Speech perception as the mapping of the speech signal onto linguistic units.

Typically, linguistic proposals that model the role of speech perception in phonology refer to the perceptual mapping of the speech signal as an extra-linguistic, general auditory and universal phenomenon.¹ Within this view, speech perception plays a role in shaping phonological systems but it is not modelled within the linguistic knowledge of language-specific sound structure. For instance, Hyman (2001: 145) argues that speakers do not need to ‘know’ the universal phonetics involved in speech perception because no evidence is available for phonology being stored in phonetic terms.² Likewise, Hume & Johnson (2001)’s model refers to speech perception as an ‘external force’ whose elements are tied up with the transduction of speech sounds in the auditory periphery. These authors claim that it would be erroneous to directly incorporate speech perception into phonological theory because it would imply that perception is exclusive to language (p. 14). Similarly, Steriade (2001: 236) proposes an external or extra-linguistic *perceptibility map* (P-map) to formalize the universal perceptual similarity constraints that have an effect on phonological phenomena, such as place assimilation. Finally, Brown (1998) refers to the phonetic mapping of the speech signal as a universal and general auditory phenomenon and, consequently, assumes that it occurs automatically and needs no phonological explanation.

Contrary to the above views, I argue that speech perception is a language-specific phenomenon that involves linguistic knowledge. This claim is not at all new to most phoneticians and psycholinguists and it is supported by a large body of empirical evidence. Cross-linguistic studies (cf. Strange 1995) have shown, for instance, that experience with the fine-grained acoustics of a specific language environment shapes listeners’ perception of the speech signal in a linguistic way. This environmental dependence is observed in two of the basic properties of speech perception, namely the categorization of acoustic continua and the perceptual integration of multiple acoustic dimensions. These properties of speech perception

¹ Here the idea that speech perception is a ‘universal phenomenon’ is taken to mean that human listeners perceive speech sounds in exactly the same manner because humans share the same physiology.

² As will be mentioned in the following paragraphs, empirical evidence suggests that phonetic properties are produced and heard differently depending on the speaker/listener’s linguistic background. In the next section, it will be proposed that a phonological grammar which contains phonetic terms or phonetic values is needed to account for the knowledge underlying the categorization of sounds, which is a phonological phenomenon because it is specific to each language or language variety.

have been shown to differ cross-linguistically (cf. Gottfried & Beddor 1998, Escudero & Polka 2003) and, even, cross-dialectally (cf. Miller & Grosjean 1997, Escudero 2001, Escudero & Boersma 2001/2003, 2004).

Furthermore, infant speech perception studies have shown that, within their first year of life, babies develop speech perception capabilities that are appropriate for their specific language environment exclusively (cf. Werker & Tees 1984; Jusczyk, Cutler & Redantz 1993; Polka & Werker 1994). This finding leads Kuhl (2000) to argue that infants develop from universal acoustic discrimination to *filtered* or *warped* language-specific perception. This means that the language-specific filtering or mapping of speech input alters the universal acoustic dimensions of the speech signal in order to highlight differences between the categories of our native language. Moreover, Kuhl claims that “no speaker of any language perceives acoustic reality; in each case, (speech) perception is altered in the service of language” (2000: 11852). However, this altering of the perceptual space seems to apply to speech only because, as shown by phonetic and psycholinguistic studies, speech perception involves different means and processes than those required by the perception of other auditory stimuli (cf. Miyawaki et al. 1975, Werker & Logan 1985, Jaquemot et al. 2003).

I interpret the evidence as supporting the claim that speech perception is not *solely* performed by our general auditory system but also by perceptual mappings that are language specific and exclusively appropriate for the language (or languages) that we have learned. Consequently, the perceptual mapping of the speech signal should be modelled within phonological theory. This line of thinking has recently been followed by a number of phonologists to model different aspects of speech perception, e.g. Boersma 1998, Tesar & Smolensky 2000, Broselow to appear, Pater 2004. Importantly, the L2 proposal that I will discuss in this paper belongs to the tradition of modelling sound perception with Stochastic Optimality Theory started by Boersma (1998), which was extended to L1 and L2 acquisition by Escudero and Boersma (2003, 2004) and Boersma, Escudero and Hayes (2003). Before moving onto the proposed model, the next section discusses a number of criteria that a comprehensive model of sound perception should incorporate, in light of the available phonetic and psycholinguistic empirical evidence.

1.2 Criteria for a comprehensive model of sound perception

A comprehensive model of sound perception needs to consider (a) the *definition* of this phenomenon, i.e. what we mean by the mapping of the speech signal, (b) the *type of process* involved, i.e. is it universal or language-specific, (c) the *elements* involved in this processing, i.e. representations, processing mechanisms or a combination of these, and (d) the *relationship between the elements* assumed to be involved in speech perception. Thus, with respect to (a), the definition of speech perception, we assume that it refers to the decoding of the variable and continuous acoustic dimensions of the speech signal. Concerning (b), the type of process involved, it is proposed here that speech perception is a linguistic and language-dependent procedure, i.e. during language development the specific language environment shapes the decoding of the speech signal. Concerning (c), the elements involved in the processing of the signal, it is proposed here that speech perception involves both abstract representations and perceptual mappings. Finally, concerning (d), it is proposed that the degree of abstraction of sound representations depends on the acoustic properties of the

signal and the way in which these properties are encoded in the perceptual mappings. I argue that all of these aspects contribute to an adequate modelling of speech perception. Thus, Table 1 shows possible ways of modelling the four essential aspects of sound perception proposed here.

Sound perception	Proposed modelling of sound perception
Definition: Decoding of acoustics	Phonetic-to-phonological mappings
Type of process: Language-specific and language-dependant	Linguistic knowledge underlies speech perception: Grammatical rules or grammatical constraints
Elements: Speech signal + mappings + abstract categories	Perceptual mappings connect the signal with the listener's abstract representations
Relationship between elements: The nature of categories depends on the signal and the mappings	The input generates the mappings and they, in turn, generate sound representations.

Table 1: Notions associated to speech perception and criteria for their modelling.

The modelling possibilities shown in Table 1 would integrate phonetic and phonological approaches to sound perception. However, a comprehensive model of sound perception would also need to incorporate two psycholinguistic constructs, viz., the pre-lexical and bottom-up nature of speech perception. That is, psycholinguistic research has shown that the decoding of the acoustic properties of the speech signal precedes the access of meaning (cf. McQueen 2005). In addition, it has been shown that speech perception takes place without the aid of word knowledge (cf. Miller & Dexter 1988, Schachter & Church 1992, Pitt & McQueen 1998, Burki-Cohen et al. 2001, Dupoux et al. 2001). Crucially, the phonological framework that will be discussed in the next section incorporates all of the criteria suggested here.

2 Linguistic Perception

The explicit modelling of speech perception as linguistic knowledge started with Boersma (1998)'s Optimality Theoretic (OT, Prince & Smolensky 1993) *perception grammar*, which was proposed to underlie the perception of the sounds of a language. This perception grammar implements the *optimal perception hypothesis*, which states that an *optimal* listener will construct those vowels and consonants that are most likely to have been intended by the speaker. Escudero & Boersma (2003)'s proposal extends Boersma's perception grammar by introducing *auditory-to-segment cue constraints* which explain how multiple auditory cues are perceptually integrated in order to map the speech signal onto phonological categories like place of articulation or voicing.

In this paper, I refer to the combination of Boersma's pioneering work on perception grammars and Escudero & Boersma's extension as the *Linguistic Perception* (LP) model, a term that was first used in Escudero (2005). This model is illustrated in Figure 2.

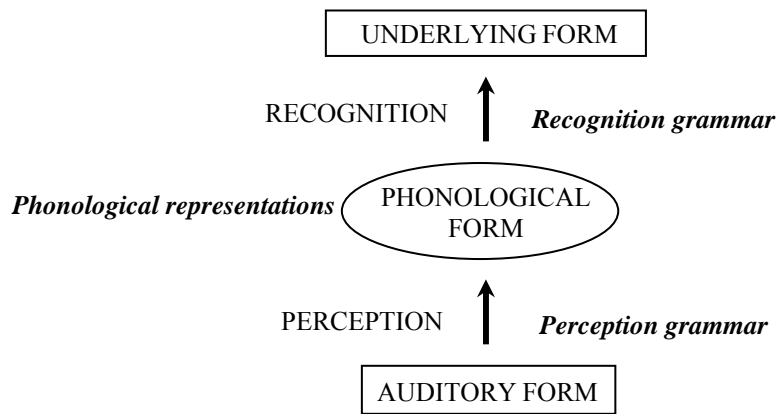


Fig. 2: A model for speech comprehension composed of two sequential mappings, perception and recognition, and three representations, auditory, phonological and underlying forms.

Thus, the LP model makes two main assumptions. First, it assumes that speech comprehension is a two-step process that involves two mappings, namely speech perception and speech recognition. Second, it assumes that speech perception is a pre-lexical and bottom-up process. These two assumptions are also found in the psycholinguistic models mentioned in § 1.2. In addition, the model incorporates the idea that both speech perception and recognition are handled by linguistic grammars, depicted as single arrows in the figure.

2.1 Optimal Linguistic Perception

Within the LP model, it is proposed that the auditory-to-abstract mapping of the speech signal depends on the specific properties of the language environment involved. Specifically, Escudero & Boersma (2003) argue that, for speech perception, listeners integrate the different auditory dimensions that they hear in ways that resemble the manner in which such dimensions are combined in speech production. This claim was formalized as the *optimal perception hypothesis*, which states that an optimal listener will prefer auditory dimensions that reliably differentiate sounds in the production of her language. In addition, such an optimal listener will identify auditory inputs as the vowels or consonants that are most likely to have been intended by the speaker. Therefore, an important prediction of this hypothesis is that differences in the productions of two languages or language varieties will lead to differences in their respective optimal perception of these two languages or language varieties. Specifically, if two languages differ in the way acoustic dimensions are used and integrated in production, the optimal listeners of these languages will have different ways of perceiving these languages. For instance, Escudero & Boersma (2003) found that Southern British English and Scottish English speakers used F1 and duration differently when producing the vowels /i/ and /I/, and they hypothesized that the optimal perception of these two languages would exhibit differences in line with the attested production differences. The data reported in Escudero (2001) showed that their hypothesis was borne out. It was shown that Scottish English listeners relied almost exclusively on F1 differences to categorize tokens of /i/ and /I/, while Southern British English listeners used both F1 and duration differences to categorize the same stimuli. Thus, this perceptual difference between language

varieties closely resembles the different ways in which acoustic dimensions are used in production.

A series of studies was recently conducted to examine whether optimal perception also applied to other cases. And indeed, Escudero & Polka (2003) found large differences between the productions of the vowels /æ/ and /ɛ/ in Canadian English (CE) and Canadian French (CF), which led to the same large differences in the perception of these vowels by native listeners of the two languages. In their production experiment, Escudero & Polka recorded the /æ/ and /ɛ/ productions of 6 (3 male, 3 female) monolingual CF speakers and 6 (3 male, 3 female) monolingual CE speakers. The vowels were produced in five different Consonant-Vowel-Consonant (CVC) environments and embedded in a carrier French or English sentence. Figure 3 shows the F1 and duration values of the 60 tokens produced in each language.

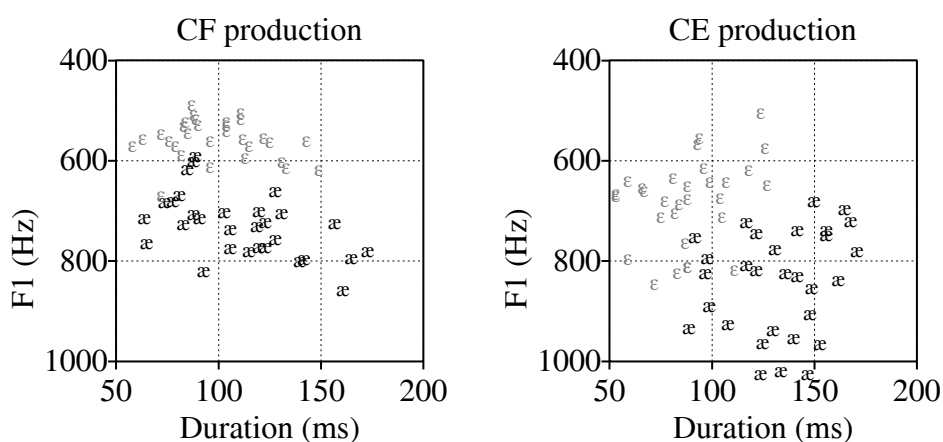


Fig. 3: F1 and duration values of the 60 CE and 60 CF tokens.

We can see that although the productions of the vowels occupy similar acoustic regions, it is clear that the use of F1 and duration is rather different. Thus, we can easily observe that while the CE vowels (on the right) are produced with dissimilar F1 and duration values, the CF vowels (on the left) are produced with different F1 values only. That is, intended CE /æ/ tokens usually have a high F1 value together with a long duration, while intended CE /ɛ/ tokens have a lower F1 value together with a short duration. For instance, the great majority of vowel tokens produced with an F1 value of 700 Hz are intended as CE /æ/ if they are longer than 110 ms but as CE /ɛ/ if shorter than 110 ms. By contrast, although intended CF /æ/ tokens also have higher F1 values than intended CF /ɛ/ tokens, the two CF vowels freely vary between long and short values. For instance, CF tokens produced with values around 700 Hz are almost always intended as /æ/ and almost never as /ɛ/, regardless of their duration values. In addition, Figure 4 shows that the average productions, represented by the symbols, and the variation between speakers, represented by the ellipses, differ across languages. Note that the dotted curve represents the *equal-likelihood* line as computed in Escudero & Boersma (2004a), which is the line along which tokens are likely to be intended as either of the two vowels.

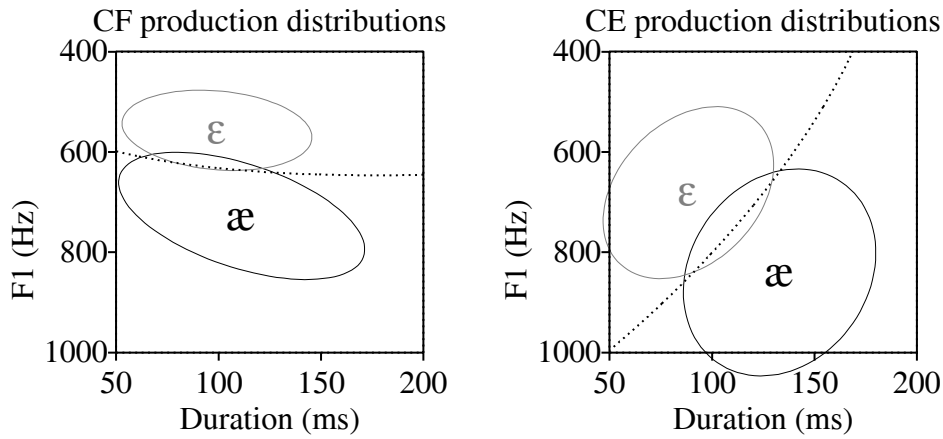


Fig. 4: CE and CF average and distributions for /æ/ and /ɛ/. Dotted curve: Equal-likelihood.

In these average production distributions, we can again observe that the acoustic dimensions are combined in a language-specific way because the directions of the ellipses in the two plots are different. That is, the ellipses for the CF vowels are almost horizontal, which means that CF speakers almost exclusively use the F1 dimension to distinguish between the two vowels. In contrast, the CE vowels have a completely diagonal shape, which means that CE speakers use a combination of F1 and duration when distinguishing the two vowels. Along with these dissimilarities in the integration of acoustic dimensions, the two languages exhibit differences in F1 distributions for the production of the two vowels. That is, the mean F1 productions of CE /æ/ and /ɛ/, viz. 840 and 681 Hz respectively, are higher than their CF counterparts, which have F1 values of 728 and 557 Hz respectively.

According to the optimal perception hypothesis, these differences in vowel production should lead to differences in perception. This is because the optimal perception of a language should resemble its production distributions and therefore the equal-likelihood line in production should resemble the perceptual category boundary. Thus, it was predicted that an optimal CF listener would rely almost exclusively on F1 and hardly on duration when distinguishing between /æ/ and /ɛ/, whereas an optimal CE listener would rely on both duration and F1 for categorizing the same vowels. Likewise, the differences in F1 distributions should lead to differential categorization of the same vowel tokens. Figure 5 shows that the optimal *perceptual category boundary* of CE and CF listeners, which is represented as a solid line, coincides with the respective *production equal-likelihood* line in Figure 4. As an example of language-specific optimal perception, we observe that the same token [785 Hz, 85 ms], which is depicted by a diamond, is perceived as /æ/ in CF but as /ɛ/ in CE.

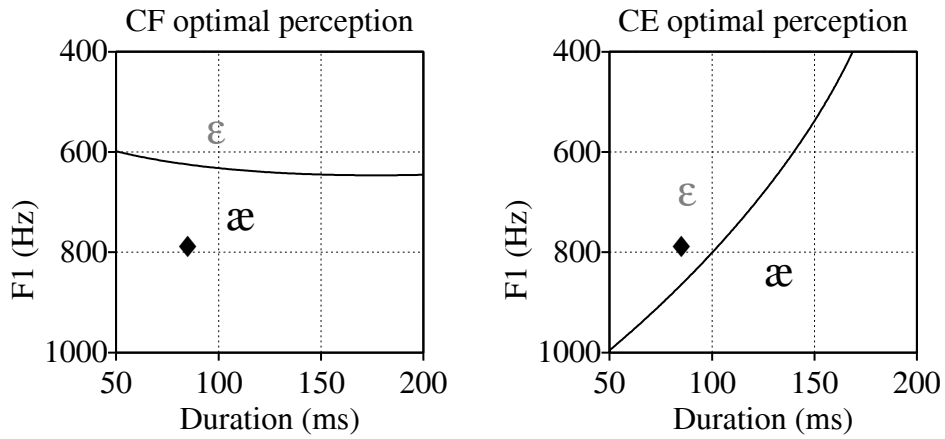


Fig. 5: Optimal CE and CF categorization of [785 Hz, 85 ms].

This optimal perception hypothesis needed to be validated with the perception of *real* CE and CF listeners. Thus, Escudero & Polka (2003) tested the perception of the 60 CE and the 60 CF vowel tokens by eight monolingual CE listeners and eight monolingual CF listeners. These listeners performed a native vowel categorization test. They were told that all of the stimuli were from their native language and were asked to choose between five of their native vowels. Thus, the CE listeners' options were *see* [si:], *it* [ɪt], *say* [seɪ], *pet* [pɛt], and *at* [æt], while the CF listeners' options were *bise* [biz], *biss* [bɪs], *bess* [bɛs], *be* [be], and *bace* [bæs].³ Figure 6 shows the perceptual category boundary for the two groups as computed in Escudero & Boersma (2004a).

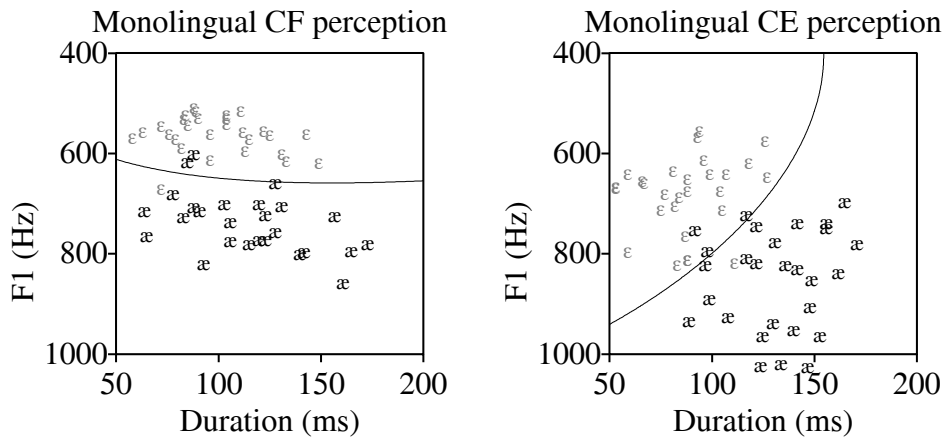


Fig. 6: Monolingual perception of the 60 CE and the 60 CF tokens.

We can see that the CE and CF listeners perceived the productions of their respective native language in a way that matches their predicted optimal perception. Thus, when comparing Figures 4, 5 and 6, the perceptual category boundaries of real listeners resemble their language-specific equal-likelihood line in production in Figure 4 and their optimal category boundaries in Figure 5. As for the mechanism that leads to this perceptual behaviour, the LP

³ The majority of the French category responses were non-sense words whose pronunciations undoubtedly led to the expected vowels.

model presupposes that linguistic knowledge in the form of a perception grammar underlies the attested language-specific optimal perception. Such an adult perception grammar can be formalized in OT by means of the cue constraints in (1), formulated by Escudero & Boersma (2003: 77-78).

(1) *Cue constraints* for adult sound perception

“A value x on the continuum f should not be mapped to the phonological category y ”⁴

- e.g. [F1=260 Hz] *→ /i/
- [F1=500 Hz] *→ /e/
- [duration=120 ms] *→ /i/
- [duration=60 ms] *→ /e/
- [F2=2800 Hz] *→ /i/
- [F2=1400 Hz] *→ /e/, ...

and so on for every duration, F1, F2 value and for every vowel.

According to the model, the constraints are ranked with respect to their distance from the centres of the production distributions of the vowels. In the case of the CF grammar, the mean F1 and standard deviation (s.d.) for /æ/ and /ε/ are 748 Hz (60) and 557 Hz (40) respectively. Thus, for a token with values [785 Hz, 85 ms], like the diamond in Figure 5, the constraint that says ‘do not perceive 785 Hz as /ε/’ will be higher ranked than the one that says ‘do not perceive 785 Hz as /æ/’ because 785 Hz is 5.7 s.d. away from the mean F1 value for /ε/ but only 0.61 s.d. from the mean F1 for /æ/. Regarding vowel duration, 85 ms is 0.87 s.d. away from the mean duration value for /æ/ and 0.61 from that of /ε/. Tableau 1 shows the *distance-based ranking* of the constraints in the CF perception grammar. Given this ranking, an optimal CF listener will perceive [785 Hz, 85 ms] as CF /æ/ because a token with such a F1 value is highly unlikely to have been intended as /ε/ in the CF production environment.⁵

[785 Hz, 85 ms]	785 Hz not /ε/	85 ms not /æ/	85 ms not /ε/	785 Hz not /æ/
/bæk/	*!		*	
☞ /bæc/		*		*

Tableau 1. The constraints and constraint rankings relevant for the categorization of [785 Hz, 85 ms] in the optimal CF perception grammar.

⁴ These constraints are different from those proposed by Boersma (1998:163-172) because the latter evaluate a mapping from values of a certain auditory continuum to other values on the same continuum. Boersma’s constraints can be called *auditory mapping* constraints and can be said to be relevant for infant perception, as will be described in § 2.2. Importantly, Escudero & Boersma (2004b) show that these auditory constraints would have trouble handling the integration of multiple cues in L1 and L2 sound perception.

⁵ The dotted line in the tableau shows that the constraints “do not perceive 785 Hz as /æ/” and “do not perceive 85 ms as /ε/” are ranked at the same height because such F1 and duration values are equally distant from the mean values of the respective vowel. However, because the highest ranked constraint rules out /ε/ as a candidate, this duration constraint can no longer play a role in the categorization of this token, as depicted by the gray shading of the column. Thus, a vowel token with a high F1 value such as 785 Hz will always be categorized as CE /æ/, irrespective of its duration value.

Tableau 2 shows that the same four constraints have a different ranking in the CE optimal perception grammar. This is because the mean F1 and duration distributions of the CE vowels (F1 /æ/= 840 Hz (103), F1 /ɛ/= 681 (86), duration /æ/= 133 (23) ms., and duration /ɛ/= 88 (21) ms., are very different from those of the CF vowels. Thus, 85 ms is 2.1 s.d. away from the mean duration value for /ɛ/ but only 0.14 away from that of /æ/, while 785 Hz is 1.2 s.d. away from the mean F1 value for /ɛ/ and 0.53 s.d. away from that of /æ/. Consequently, the resulting distance-based ranking shown in Tableau 2 has the constraint “do not perceive 85 ms as /æ/” as the highest ranked, because a token with such duration value is very unlikely to have been intended as the vowel /æ/ in the CE production environment. Note that this ranking is the only possible one if we follow the model’s ranking proposal, which is based on the distances between auditory values. Thus, in the optimal CE grammar, duration constraints play an important role in determining the perceived vowel category, a situation which contrasts with that of the optimal CF perception grammar shown in Tableau 2. As a consequence, a token with short vowel duration such as 85 ms is categorized as /ɛ/, irrespective of its F1 value. This vowel categorization is different from the one resulting from the CF perception grammar. It is proposed here that the difference in the constraint rankings of the two perception grammars underlies the cross-linguistic perceptual difference shown in Figure 6.


[785 Hz, 85 ms]	85 ms not /æ/	785 Hz not /ɛ/	785 Hz not /æ/	85 ms not /ɛ/
 /bɛk/		*		*
/bæk/	*!		*	

Tableau 2: The constraints and constraint rankings relevant for the categorization of [785 Hz, 85 ms] in the optimal CE perception grammar.

In sum, the LP model is able to account successfully for the way in which real adult listeners perceptually map the auditory dimensions of the speech signal onto sound categories. Perhaps more importantly, the model is able to formalize the linguistic knowledge that underlies the attested perceptual behaviour by means of a phonological implementation of the optimal perception hypothesis.

2.2 L1 acquisition of Linguistic Perception

The next question in the modelling of sound perception is how adult listeners attain optimal perception. In answering this question, Boersma, Escudero & Hayes (2003) put forward an account of the path that an infant follows when learning to perceive sound categories. They proposed that the three auditory-mapping constraint families described in (2) are first introduced in the infant perception grammar. These three types of constraints are similar to Boersma’s perception grammar constraints (Boersma 1998).

(2) Auditory mapping constraints in the infant perception grammar

PERCEIVE ($f: x$)

‘Map the value x along the continuum f to some value along that same continuum’

*CATEGORIZE ($f: y$)

‘Do not perceive anything as the value y along the continuum f ’

*WARP ($f: d$)

‘Do not perceive a value along a continuum f as a value that is a distance d (or more) away along that same continuum’.

As we can see, PERCEIVE constraints allow the infant to map auditory tokens onto perceived counterparts. On the other hand, *CATEG constraints forbid the classification of linguistic input, and *WARP constraints impede their modification. With respect to the initial ranking of these continuous auditory constraints, Boersma, Escudero & Hayes (2003) propose that all *CATEG constraints are ranked higher than the PERCEIVE constraints. This means that, initially, an infant cannot map the input onto any category. In addition, *CATEG constraints are ranked higher than the *WARP constraints that do not change the identity of the input in an auditorily noticeable way. Kewley-Port (1995) found that the just noticeable difference (JND) between F1 values is 40 Hz for adult listeners. Thus, *WARP constraints with a value of 40 Hz or below are ranked very low because they lead to a very small change between the input and the perceived category. To exemplify the workings of the three constraints families, Tableau 3 shows the initial perception grammar for a CE infant that is confronted with an F1 value of 670 Hz, which is a common F1 value for the CE /ε/.

[670 Hz]	*CATEG (/630/)	*CATEG (/670/)	PERCEIVE ([670])	*WARP (40)
/630 Hz/	*!			*
/670 Hz/		*!		
☞ /-/			*	

Tableau 3: The null perception at the initial state in learning to perceive sound categories.

Here we can see that, at the initial state, the infant may not manage to perceive the input as any category, perhaps because she does not hear it as linguistic yet, in which case the input will result in a *null perception*. Therefore, some perceptual development needs to occur to allow the infant to be able to categorize the sounds of her language. Boersma, Escudero & Hayes employ the Gradual Learning Algorithm (GLA, Boersma & Hayes 2001) as the learning device responsible for such development. Thus, the GLA initially performs the classification of language sounds via their auditory distributions, a mechanism that can be called *auditory-driven perceptual learning*. In order to avoid the null perception of Tableau 3, the infant’s GLA first acts as an *identity matching* device which will *force* the grammar to classify any auditory input as its perceived counterpart. That is, the GLA will tell the infant that she should have categorized [670 Hz] as /670/. As a result, the constraint *CATEG /670/ will be demoted in order to increase the match between input and output. This identity matching mechanism is demonstrated in Tableau 4. Here we see that, because of the

command given by her GLA, the infant automatically realizes that the null perception is incorrect, as depicted by the asterisks, and that she should have classified the auditory input as its perceived counterpart, as depicted by the check mark.

[670 Hz]	*CATEG (/630/)	*CATEG (/670/)	PERCEIVE ([670])	*WARP (40)
/630 Hz/	*!			
✓ /670 Hz/		*!→		*
☞	/-/		←*	

Tableau 4: GLA identity matching procedure.

As we can see in the example, the identity matching procedure performed by the GLA will result in the lowering of the constraint with the value of the incoming auditory event, in this case 670 Hz, and in the rising of the PERCEIVE constraint for this same value. Consequently, the next time that the infant hears [670 Hz] it will be more likely that she perceives it as /670 Hz/. It is important to mention that although the infant will hear a large number of auditory inputs, not all *CATEG constraints will be demoted equally fast because the auditory values with which language sounds are produced commonly have particular frequency distributions. Consider, for instance, how F1 values are distributed in the production of CE /æ/ and /ɛ/. Figure 7 shows idealized F1 distributions of the CE vowels, assuming that they form Gaussian shapes with a standard deviation of 0.166 octaves.

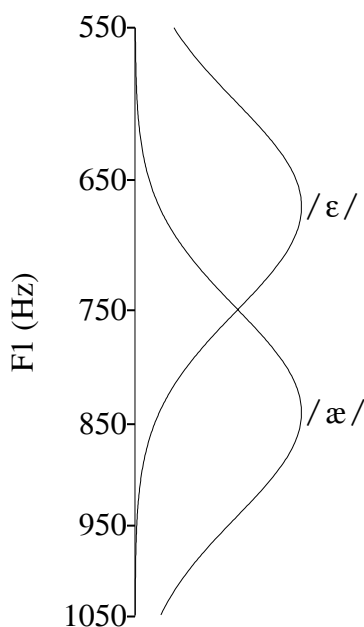


Fig. 7: F1 distributions of CE /æ/ and /ɛ/.

We can observe that the peaks of the two Gaussian curves lie at 670 and 840 Hz. These two values represent the most common tokens for the vowels / ϵ / and / \ae / respectively. Boersma, Escudero and Hayes further propose that GLA auditory-driven learning changes the infant’s perception grammar to appropriately cope with the distributional properties of her production environment. Tableau 5 illustrates how the infant perception grammar handles a rather uncommon F1 value, e.g. [710 Hz], in the vowel productions of CE speakers.

[710 Hz]	*CATEG (/750/)	*CATEG (/710/)	*CATEG (/670/)	*WARP (40)
☞ /670 Hz/			*	*
/710 Hz/		*		*
/750 Hz/	*!			

Tableau 5: Infant perception after some identity matching learning.

As we can see, the high frequency of 670 Hz in the CE environment has resulted in the low ranking of the constraint that limits its categorization as /670/. In contrast, because of the uncommon nature of auditory values such as 710 Hz and 750 Hz, the constraints that ban their perceived counterparts are high ranked. As a result, the infant’s grammar will choose the most frequent candidate because of the low ranking of the corresponding *CATEG constraint. In the example of Tableau 5 we only see a simplified list of perceived candidates due to space limitations but it is of importance to bear in mind that any F1 value is a potential perceived candidate during auditory-driven learning.

Furthermore, Boersma, Escudero & Hayes (2003) explain how a reiteration of *frequency-driven categorization* leads to the mapping of several auditory inputs onto the most frequently perceived categories. That is, the infant’s GLA adjusts the perception grammar in such a way that auditory values will be mapped onto a finite number of auditory categories, i.e. the most frequent ones. Crucially, this finite set of categories will automatically be turned into more abstract categories corresponding to the ones produced in the infant’s environment. In the case of the CE infant, the two categories that result from the auditory-driven learning of the F1 distributions in Figure 7 can be called /mid-low/ and /low/.⁶ This GLA auditory-learning is compatible with the recent findings that suggest that infants are able to calculate the statistical distributions of the auditory values of sound productions and that this ability leads to the creation of phonetic categories (cf. Maye, Werker & Gerken, 2002). Furthermore, it leads to the same warping of the infant perceptual space that has been shown to occur in the first year of life (Kuhl 1991).

⁶ Although this paper only discusses the perception of two categories and not that of a system of vowel sounds, we use vowel height phonological features to refer to the categories present in the baby’s lexicon. This is because the model proposes that sound categorization is uni-dimensional in early stages of acquisition and that the integration of dimensions, such as F1, F2 and duration to form vowels such as / \ae / and / ϵ / only occurs later in life, as will be mentioned in the next paragraph. In addition, the model proposes that adult sound categorization is characterized by the integration of multiple cues for the perception of sounds and therefore the modelling of adult L1 and L2 categorization employs vowel segments as the candidates of an adult perception grammar.

With respect to further perceptual development, Boersma, Escudero & Hayes (2003) propose that once an abstract lexicon is in place the infant’s GLA can re-rank the mapping constraints in the perception grammar when faced with mismatches between perceived and lexicalized representations. This second mechanism is known as lexicon-driven learning because the abstract lexical representations trigger re-rankings in the perception grammar, leading to optimal perception. For instance, imagine that a [785 Hz] production is intended as a /mid-low/ vowel but the child, by mistake, perceives it as a /low/ vowel, as shown in Tableau 6.

[785 Hz] <i>/b-mid vowel-k/</i>	785 Hz not /mid-low/	785 Hz not /low/
✓ /mid-low/	*!→	
✗ /low/		←*

Tableau 6: GLA lexicon-driven learning.

At this point the baby has access to the semantic context of the words she hears, which in this case reveals that the speaker intended a word containing a /mid-low/ vowel, as shown in the input to the grammar. Therefore, the child will automatically notice that in this context the correct perception should have been /mid-low/, as depicted by the check mark.. When confronted with this situation, the child’s GLA re-ranks the constraints in the perception grammar so as to enable the perception of the next [785 Hz] token as /low/. Specifically, this is achieved by lowering the constraint against perceiving the acoustic value as /low/ and by simultaneously raising the one against perceiving the same value as /mid-low/, a procedure that was first proposed and described in Boersma (1998). This second type of perceptual learning results in the category boundary shifts which have been shown to occur developmentally in infants and children (cf. Nittrouer & Miller 1997; Gerrits 2001; Jones 2003). The question for now is whether the same two types of learning mechanisms are present in adult L2 acquisition. The remainder of this article presents a phonological model for explaining the development of sound perception in second languages. This L2 model is based on the Linguistic Perception framework for sound perception and L1 acquisition discussed above.

3 L2 Linguistic Perception of SIMILAR L2 sounds

The Second-Language Linguistic Perception (L2LP) model aims at describing, explaining, and predicting L2 sound perception at the initial, developmental, and end states. In this paper, the model will be applied to the acquisition of L2 sounds that are phonemically equivalent but phonetically different from L1 sounds, which are also called SIMILAR L2 sounds. For this L2 scenario, the model predicts that learners will equate two L2 phonemes with two L1 phonemes for purposes of lexical storage, as shown on the left of Figure 8. This lexical equation has an origin in perception because many tokens of the L2 sounds have auditory properties which are similar to the properties of the corresponding L1 sounds. However, this scenario also features a mismatch in the mapping from auditory events to phonological categories. This is because some tokens or phonetic realizations, commonly written between

‘[]’, of the L2 categories are unlikely to be perceived as their L1 phonological counterparts, as depicted by the thick lines in Figure 8, right.

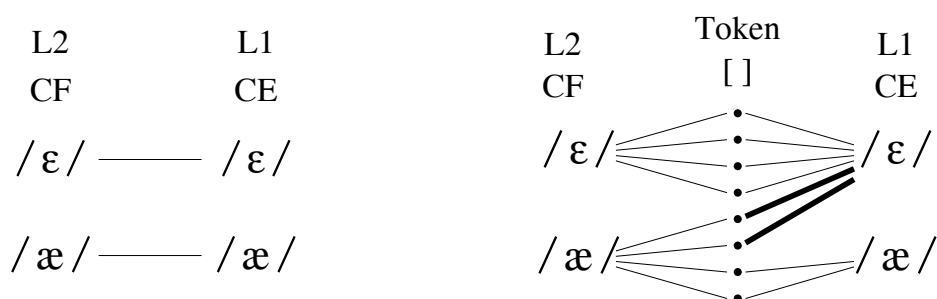


Fig. 8: Phonemic equation and perceptual mapping in the SIMILAR L2 perception scenario.

According to the L2LP model, if two L2 sounds are equated to a single sound in the L1, the learner faces the common NEW sounds scenario. But if two L2 sounds are equated to two L1 sounds, the learner faces a SIMILAR scenario. This differentiation between “new” and “similar” sounds is also found in many other models of L2 sound perception, such as Flege’s *Speech Learning Model* (Flege 1995, 2003, 2004), Best’s *Perceptual Assimilation model* (Best 1995, 2003), Major’s *Ontogeny Phylogeny Model* (Major 2001, 2002), Kuhl’s *Native Language Magnet model* (Kuhl 2000), and Brown’s *Phonological Interference model* (Brown 1998, 2000). These proposals make diametrically opposite claims regarding L2 similar sounds, namely they either suggest that this scenario poses *no* L2 learning challenge (Brown, Best, and Kuhl), or that it poses *the greatest* L2 challenge (Major, and Flege). The authors that follow the first approach share the idea that the presence of L2 sounds, features, or phonetic dimensions in the L1 guarantees the absence of an L2 perceptual learning problem, and therefore a SIMILAR scenario does not pose a challenge to either the L2 learner or the researcher. In contrast, the authors that follow the second approach claim that SIMILAR L2 sounds are the most difficult to acquire because the L2 learner will not be able to master them without an effect on their L1. For instance, Flege claims that SIMILAR L2 sounds will be equated to L1 sounds and therefore L2 learners will not be able to form new L2 categories for these sounds, which in turn results in non-native perception.

Thus, different approaches to L2 sound perception assume different and even opposite L2 tasks in a SIMILAR scenario. That is, one approach assumes that L2 learners will have *no* task when learning this type of sounds because of the assumption that having identical categories in L1 and L2 automatically turns the learner into a native-like perceiver. In contrast, the second approach claims that in this scenario the goal of L2 category formation is extremely difficult to achieve and learners may therefore never be able to attain full L2 proficiency when confronted with SIMILAR L2 sounds.

The L2LP model (cf. Escudero & Boersma 2004b and Escudero 2005) that is presented in this paper proposes an alternative approach to the development of SIMILAR L2 sounds. Unlike the first approach to the phenomenon, the L2LP proposes that SIMILAR sounds *do* pose a learning challenge, namely the adjustment of perceptual mappings. In addition, unlike the second approach, the L2LP claims that learners faced with this scenario needs to adjust their existing L1 categories instead of creating new L2 categories. Thus, the L2LP model, also

unlike the second approach, claims that SIMILAR L2 sounds are easier to master than L2 sounds that do not exist in the learners' L1, i.e. NEW sounds. Table 2 shows the initial state, learning tasks and degree of difficulty that are hypothesized in the L2LP model.

L2LP proposal	Prediction for NEW	Prediction for SIMILAR
Initial state	Too few categories	Same number of categories
Perceptual task	1. <i>Create</i> perceptual mappings 2. <i>Integrate</i> auditory cues	Adjust perceptual mappings and category boundaries
Representational task	1. <i>Create</i> phonetic categories 2. <i>Create</i> segments	None
Degree of difficulty	Very difficult	Not difficult

Table 2: Comparative initial states and learning tasks in the NEW and SIMILAR scenarios.

As we can see, the L2LP assumes that both scenarios pose a learning challenge to the L2 learner. However, the degree of difficulty for these two scenarios is different depending on the learning tasks that need to be performed in order to attain optimal perception. Importantly, these different degrees of L2 difficulty also refer to the number of learning mechanisms involved in the two scenarios. That is, the NEW scenario, in which the learning task is to create new perceptual mappings and categories, will involve both the learning mechanisms of category creation and of boundary shifting in L1 (cf. § 2.2), while the SIMILAR scenario will *only* involve the boundary shifting mechanism. In the next sections, I show how the L2LP's principled separation between perception grammars and sound representations is used to more adequately explain the initial state, learning tasks, development and ultimate attainment in learning to perceive L2 SIMILAR sounds.

3.1 L2LP ingredient 1: Comparing the L1 and the target L2

The L2LP model proposes that the first step into explaining L2 sound perception is to describe the optimal perception of each of the languages involved. In § 2, it was mentioned that this hypothesized optimal perception was attested in the categorization of human listeners. Recall that the optimal perception hypothesis says that an optimal listener has a perception grammar that has been shaped by the acoustic properties of her production environment. Thus, if our aim is, for instance, to explain how CE listeners can learn to perceive CF /æ/ and /ɛ/, we must *first* describe the optimal perception of CE and of CF monolingual listeners.

Further, the L2LP model makes a principled distinction between perceptual mappings (performed by the perception grammar) and sound representations (constructed by the perception grammar). One of the model's main claims is that analysing mappings and categories separately results in an adequate description and explanation of the comparative knowledge underlying sound perception in listeners with different language backgrounds. For instance, the productions of the same abstract categories /æ/ and /ɛ/ exhibit different F1 distributions in CF and CE. The average F1 values for the CE vowels are 840 Hz and 681 Hz

respectively, the acoustic distance between them is $\log_2(840/681) = 0.3$ octaves, and therefore the boundary between their productions can be located at $\log_2(840) - 0.15 = 9.563$ octaves which is equivalent to 756 Hz. In contrast, the average F1 values for the CF vowels are 728 Hz and 557 Hz, the F1 distance between them is 0.39 octaves, and their boundary lies at 637 Hz. Therefore, optimal perceivers of each language will behave differently when categorizing the same set of F1 values, as shown in Figure 9. Note that the boundaries in the figure are only optimal if we assume that the input tokens only differ along the F1 dimension, i.e. if they have ambiguous duration values.

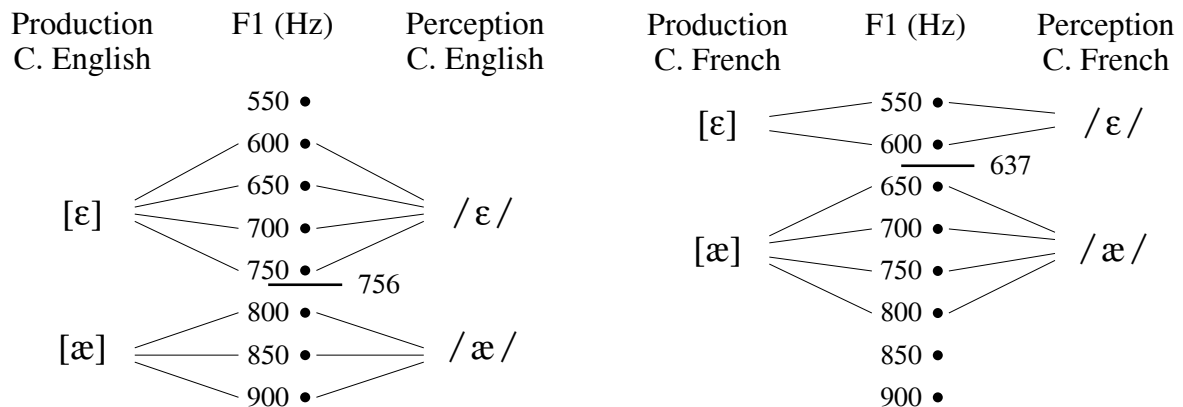


Fig. 9: CE and CF perceptual mappings and representations for /æ/ and /ε/.

We can observe that tokens between 637 and 757 Hz will be categorized as /ε/ by an optimal CE listener but as /æ/ by an optimal CF listener. Note that in both languages the tokens that have the boundary values boundaries, namely 756 Hz for CE vowel productions and 637 for CE, are the most ambiguous tokens for these two vowel categories. That is, assuming that the listeners' only choices are these two vowels, the tokens with boundary values will be categorized 50% of the time as /æ/ and 50% as /ε/ in each language. With respect to the cross-linguistic differences, tokens with values within the two languages' boundaries, i.e. between 637 Hz and 756 Hz, will be mainly categorized as /æ/ by CF listener but both as /ε/ and /æ/ by CE listeners. Thus, we can conclude that for vowel tokens with F1 values above 550 Hz, the CE and the CF optimal perception grammars may output the same two categories /æ/ and /ε/ but with different classification distributions, resulting from a difference in perceptual boundary locations.

An important prediction concerning the learning of L2 sounds is that there are two learning tasks, a perceptual and representational one. Crucially, when perceptual mappings are the only source of difference, the learner will just have a perceptual learning task. Thus, this L1 and target L2 comparison can be used to determine the initial state for the L2 learning process, at least if one assumes that L1 categories and L1 perception grammars are fully transferred to the initial state of L2 acquisition, as will be claimed in § 3.2 below. In addition, the L1 and L2 comparison allows us to determine the L2 learning tasks, the characteristics of the mechanisms underlying L2 development, and the L1 perception that the learner needs to maintain, as will be described in § 3.3-3.5.

3.2 L2LP ingredient 2: The initial state

A SIMILAR scenario will be first manifested as the equation of two L2 phonemes to the correspondent two L1 ones.⁷ According to the L2LP model, this situation arises from the automatic and unconscious *reuse* or *copy* of the L1 categories and perception grammar. This L2 initial strategy finds a linguistic formalization in the Full Copying hypothesis, which is the speech perception interpretation of Schwartz & Sprouse's (1996) Full Transfer/Full Access hypothesis.⁸ In the case of CE learners of CF, it is proposed that they will use their phonologically equivalent L1 categories /æ/ and /ɛ/ and their L1 optimal perception, i.e., their CE production distributions composed by the average productions and the optimal category boundary, to categorize CF vowels, as shown in Figure 10.

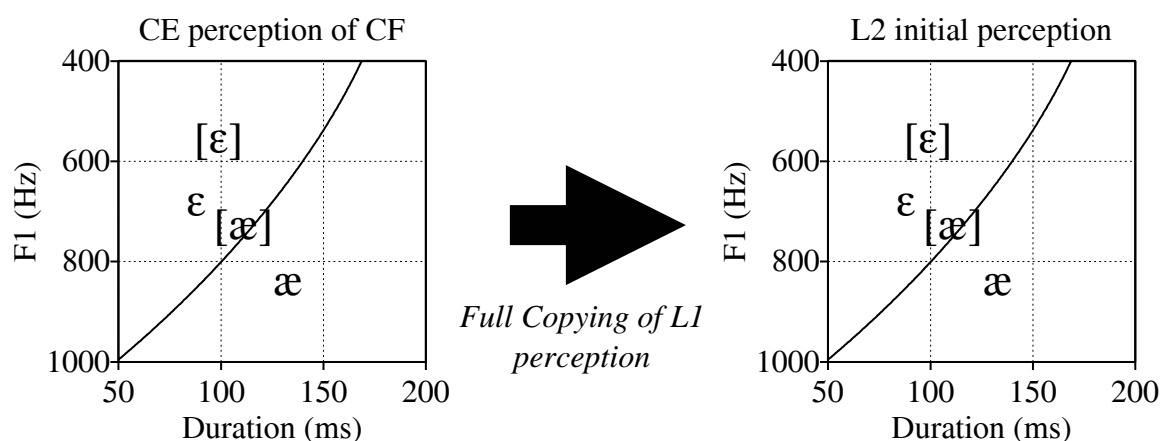


Fig. 10: Cross-language categorization of CF copied onto the L2 initial state. Between brackets: average CF productions. Curve: CE perceptual boundary.

Some support for Full Copying of L1 perception grammars and categories is shown in Escudero (2001) where Spanish learners of Scottish English had almost native-like perception of Scottish English /i/ and /ɪ/, while Spanish learners of Southern English used only duration differences to identify the same two vowels. These results can only mean that the Spanish learners copied their Spanish perception, which categorizes two Spanish vowels when confronted with Scottish English but a single Spanish vowel when confronted with Southern British English (cf. Escudero & Boersma 2004b).

⁷ It is important to acknowledge that this paper is restricted to one contrast of the L2 that is equated to another contrast in the L1. Clearly, this is a simplification of the acquisition of a vowel or consonant system in L2 acquisition. However, Escudero (2005:Ch.3) claims that every sound contrast in the target L2 could be seen as representing one of three main learning scenarios, namely NEW, SIMILAR and SUBSET. The first scenario refers to the learning of L2 sounds that do not exist in the learner's L1, a scenario, which was described in § 3. The third scenario refers to L2 sounds which already exist in the L1 but have multiple L1 correspondents. That is, in this scenario, the L1 has more categories than the L2 and therefore the L2 categories constitute a *subset* of the L1 ones, a scenario which is fully described in Escudero (2005: Ch 6). In this paper, I concentrate on SIMILAR L2 sounds which are sounds that have the same number of counterparts in the L1 but have different production distributions.

⁸ The same interpretation is found in Escudero & Boersma (2004b). Importantly, this interpretation provides the linguistic mechanism underlying Best's (1995) *two-category assimilation* and Flege's (1995) *equivalence classification* hypotheses.

Regarding empirical evidence in support of the use of L1 optimal perception to categorize foreign language stimuli, Escudero & Polka (2003) also tested the perception of CF tokens by monolingual CE listeners who were presented with the 60 CF tokens during the same perception experiment reported in § 2.1. It was predicted that CE listeners would use their L1 optimal perception to classify the CF vowels, i.e., that they would integrate duration and F1 acoustic properties when identifying vowels. Figure 11 shows Escudero & Boersma's (2004a) analysis of the findings. Note that the question marks in the figure represent the tokens that were not categorized as a single vowel by the majority of the listeners, viz. 6 out of 8.

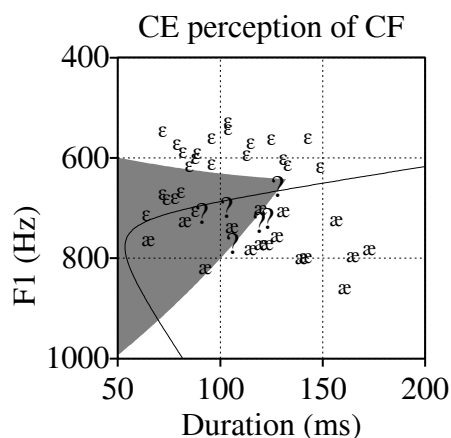


Fig. 11: Cross-language perception of CF /æ/ and /ɛ/ by CE monolingual listeners.

In the figure, we observe how eight CE monolingual listeners classify CF /æ/ and /ɛ/ tokens as their own native CE /æ/ and /ɛ/ vowel categories. The solid line in the figure is the listener's perceptual category boundary line which connects the F1 and duration values that are likely to be perceived as both vowels. This line is computed from the responses that the listeners gave to the 60 CF tokens and therefore it represents their perception of non-native vowels or their *cross-language perception*. When comparing this cross-language category boundary line to the ones shown in Figure 9, it may seem that the CE cross-language perception of the CF vowels is closer to the CF native boundary than to the CE native perceptual boundary of the listeners. However, the influence of these listeners' L1 perception is shown in the categorization of the CF tokens in the grey region.

The grey region in Figure 11 represents the area where most native tokens, i.e. English tokens for CE listeners and French tokens for CF listeners, were perceived as /æ/ by CF listeners but as /ɛ/ by CE listeners, as shown in Figure 9 above. In this figure, we observe that, when having to categorize CF tokens, the CE listeners identified most of the CF tokens which fall in the grey region as /ɛ/. This cross-language categorization pattern does not follow the native CF perception but rather the listeners' use of L1 perception strategies. That is, unlike CF native listeners, the CE listeners rely on both vowel duration and F1 to identify vowel categories, a strategy which is shown by the diagonal shape of their cross-language category boundary.

As a result, CF tokens with relatively low F1 values, viz. at approximately 700 Hz, that are produced with a short vowel duration are most likely to be identified as /ɛ/ by these CE

listeners, whereas they are categorized as /æ/ by the native CF listeners. In addition and as a result of the usage of their L1 perceptual boundary, the CE cross-language F1 boundary which falls on the grey region is 200 Hz lower than that of the native CF boundary.⁹

3.3 L2LP Ingredient 3: Predicting the L2 learning task

Given their initial state, beginning L2 learners will be able to differentiate between similar L2 sounds because these two L2 sounds 1) match two L1 categories and 2) have production distributions which overlap with the acoustic-auditory regions of two L1 categories. However, there will be a degree of mismatch between the L1 optimal perception grammar and the target language optimal perception grammar because of differences in the productions of the two categories involved. Figure 12 shows the region with the largest mismatch between the CE and CF vowel productions, which was shown for vowel perception in Figures 9 and 11 above.¹⁰

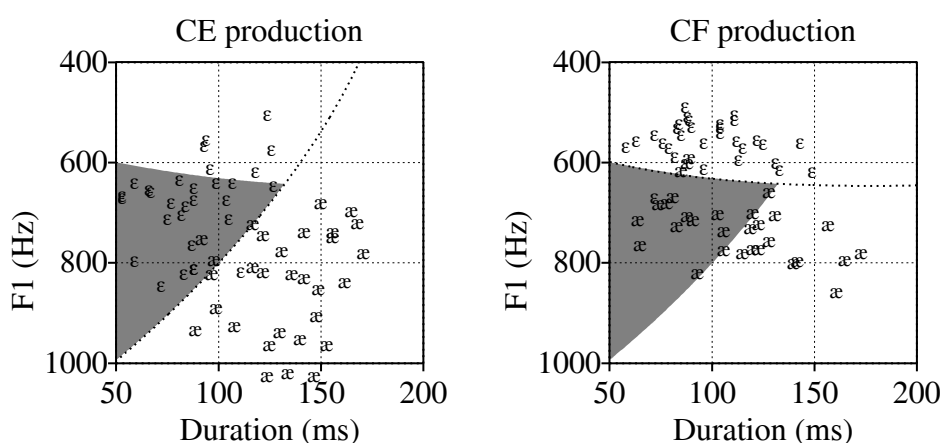


Fig. 12: Region of cross-language difference.

⁹ The monolingual CE listeners might have relied on other cues (apart from F1 and duration), such as F2 values, to categorize the CF tokens. If the listeners relied on F2 differences, a token with a low F1 value and a short duration may still yield an /æ/ native or cross-language categorization if its F2 value is too low to support /ε/ categorizations. In fact, it has been shown that when confronted with ambiguous tokens, English listeners may rely on cues that would only be secondary when categorizing unambiguous native tokens (cf. Hillenbrand et al. 2000). However, for purposes of predicting the L2 initial state and development, the two cues considered in the present article, namely F1 and duration, seemed to be extremely informative precisely because L2 learners were shown to have developmentally adjusted their perception of those cues. Furthermore, this development clearly shows that L2 re-categorization is possible and that it is performed through the adjustment of perceptual boundaries and perceptual cue weighting or trading, which is an instance of L1-like development.

¹⁰ When looking at the figures, one can also think of the top-right corner of the figure as a region of mismatch. However, only one token of the CF vowels and none of the CE vowels had values that fell on this acoustic area. Therefore, it cannot be said that there is a production mismatch in this area, and consequently it cannot be predicted that this area will constitute a problem in L2 perception because the learners may never hear tokens with such values.

As we can see, about 50% of the CF /æ/ tokens were produced in the grey region while in CE only /ɛ/ tokens can be found in that same region. Thus, we can safely assume that CE listeners will perceive up to half of the CF tokens as the other L1 category with which they have equated the CF vowels, i.e. /ɛ/.¹¹ This means that CE beginning learners of CF are predicted to categorize many tokens in that region as /ɛ/ and not as /æ/. In addition, the learners will sometimes access words in their L2 lexicon that were not intended by the speaker of the target language because an inaccurate perception will trigger access of an incorrect lexical item. Therefore, their learning task will be to adjust their initial L2 perception grammar, which is a copy of their L1 perception grammar, so as to shift their L1 boundary to the location of the L2 boundary, i.e. from the dotted line to the solid line shown in Figure 13.

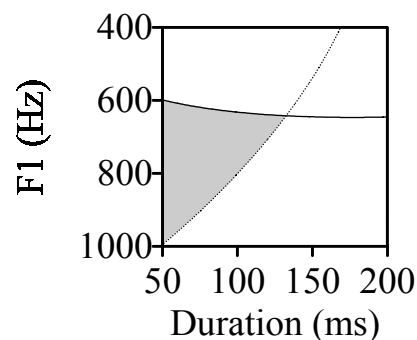


Fig. 13: Region of cross-language perceptual mismatch (in grey) and L2 learning task for CE learners of CF. Dashed line: L2 initial state. Bold line: Target L2 boundary.

As can be seen, in order for a learner to obtain native-like perception in CF, a *shift* in the category boundary needs to occur, i.e., from the dotted line to the solid line delimiting the grey region at the bottom and top respectively. This shift or adjustment, then, represents the L2 learning task because it defines the way in which the initial L2 perception grammar will need to change in order for the learner to acquire optimal L2 perception. In the case at hand, the initial L2 perception of CE learners of CF will need to change the perception of two dimensions, namely duration and F1, in order to turn their diagonal boundary, as represented by the dotted curve, into the CF optimal boundary, as represented by the solid line. That is, they must learn to ignore the durational differences between vowel tokens and to shift their F1 boundary between /æ/ and /ɛ/ to a higher location. Specifically, they need to classify the tokens in the grey region, i.e. tokens with durations shorter than 110 ms and with F1 values between 600 and 780 Hz, as their L2 /æ/ category instead of an L2 /ɛ/ category.

3.4 L2LP Ingredient 4: L2 development

The L2LP model provides a formal account of the learning mechanisms involved in the L2 learning task, an account that is based on the LP framework discussed in § 2.1 and in Escudero & Boersma (2004b). Crucially, it is proposed that L2 learners have access to the

¹¹ However, as we have seen in the previous section, the CE listeners may be able to rely on other cues to correctly categorize the CF vowels as one of their two L1 categories. If we assume that they can only rely on F1 and/or duration the prediction of 50 % incorrect categorizations still holds.

same GLA learning mechanisms available for L1 learning, namely auditory-guided category formation and lexicon-guided boundary shifting (cf. § 2.2).

In the learning of SIMILAR L2 sounds, it is predicted that L2 development will only involve a change in the perception grammar because the copied L1 abstract phonological categories are retained and remain in use for L2 lexical representation. Given that boundary shifts along different dimensions are also the result of GLA lexicon-driven learning (cf. § 2.1.2), the CE learners' task can be performed by this L1-like learning mechanism. Recall that this type of learning mechanism is activated when there is a mismatch between the perceived category and the speaker's intended word. For instance, if a beginning CE learner perceives /bɛk/ when a CF speaker produced the word *back*, GLA lexicon-driven learning will take place. That is, the semantic context will tell the learner that she should have perceived a different vowel category, namely the one contained in the word that was intended by the speaker. Thus, the learners' GLA will demote the constraints against perceiving certain F1 and duration values as the L2 /æ/ category, as shown in Tableau 7. Note that this constraint ranking is identical to the adult CE ranking shown in Tableau 2 for duration but different for F1 constraints because of the new value of 750 Hz as input to the grammar. That is, a value of 750 Hz is 0.69 standard deviations (s.d.) away from the mean F1 value for CE /ɛ/ but 0.87 s.d. away from the mean value for /æ/, and therefore this F1 value is less likely to be categorized as /æ/ than as /ɛ/ in the CE grammar:

[750 Hz, 85 ms] /bæk/	85 ms not /æ/	750 Hz not /æ/	750 Hz not /ɛ/	85 ms not /ɛ/
☞ /bɛk/			← *	← *
/bæk/	*! →	* →		

Tableau 7: Predicted lexicon-driven constraint re-ranking for Canadian English learners of Canadian French.

As proposed for L1 perceptual learning in Tableau 6 in §2.2 above,, the error in Tableau 7 will yield to the same lexical-driven perceptual learning performed by the Gradual Learning Algorithm, an adjustment which will gradually change the ranking of the constraints in the L2 perception grammar. In turn, this constraint re-ranking will lead to the two changes which could be observed in the L2 vowel categorization by CE learners of CF, namely 1) the CE learners learned to ignore the duration differences between the L2 vowels , and 2) they gradually shifted their initial F1 boundary location between the two L2 vowels. The deminishing importance of vowel duration is instantiated as a new ranking of duration-to-vowel cue constraints in the perception grammar. That is, duration constraints are now ranked in such a way that they play hardly any role in determining the winning L2 vowel category. As a consequence, the CE learner will exhibit a horizontal L2 boundary situated at a lower F1 value than her original L1 boundary. Figure 14 illustrates this gradual multidimensional boundary shift.

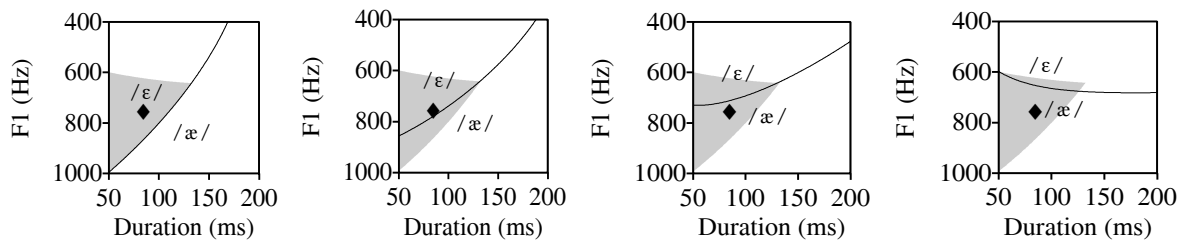


Fig. 14: Predicted category boundary shift for CE learners of CF /æ/ and /ɛ/.

In sum, it is predicted that in this scenario the learner will first equate two L2 categories to two L1 categories and therefore starts out with a near-optimal L2 perception. As a result of the mismatch between her copied L1 perception and the optimal target L2 perception, the learner will not be able to correctly categorize all L2 tokens. When faced with this situation, the learners' GLA, which in this situation acts as an error-driven constraint re-ranking mechanism triggered by mismatches between the output of perception and the lexicon, will change their perception grammars by small steps in order to decrease the probability of semantic mismatches. Finally, an optimal L2 perception will be attained when such mismatches no longer occur.

As for empirical evidence that supports this prediction, Escudero (in progress) examined the L1 and L2 perception of 21 CE learners of CF who were enrolled in a French language course at the McGill Language Centre. All learners were originally from non-French speaking regions of Canada that are outside the province of Quebec. They had monolingual Canadian English-speaking parents, and had come to Montreal at the age of 18 years. Their age at the time of testing was between 18 and 25. The learners were divided into three exposure groups, viz. beginning, intermediate, and advanced, on the basis of a language background questionnaire that determined their exposure to French in comparison with English. The target stimuli were the same 60 CF /æ/ and /ɛ/ tokens presented in the monolingual and cross-language experiments reported above, which were now presented as L2 stimuli. As for the response options, we asked the subjects to choose from five French keywords, viz. *qui* 'who', *dix* 'ten', *fait* 'do', *chez* 'at', and *ta* 'your'. The squares on the left of Figure 15 show the L2 perception for the three groups of learners while the squares on the right show the learners' L1 perception of the same CF tokens (see § 3.5). Note that the question marks in the figure represent the tokens that were not categorized as a single vowel by the majority of the learners in each group, viz. 5 out of 7.

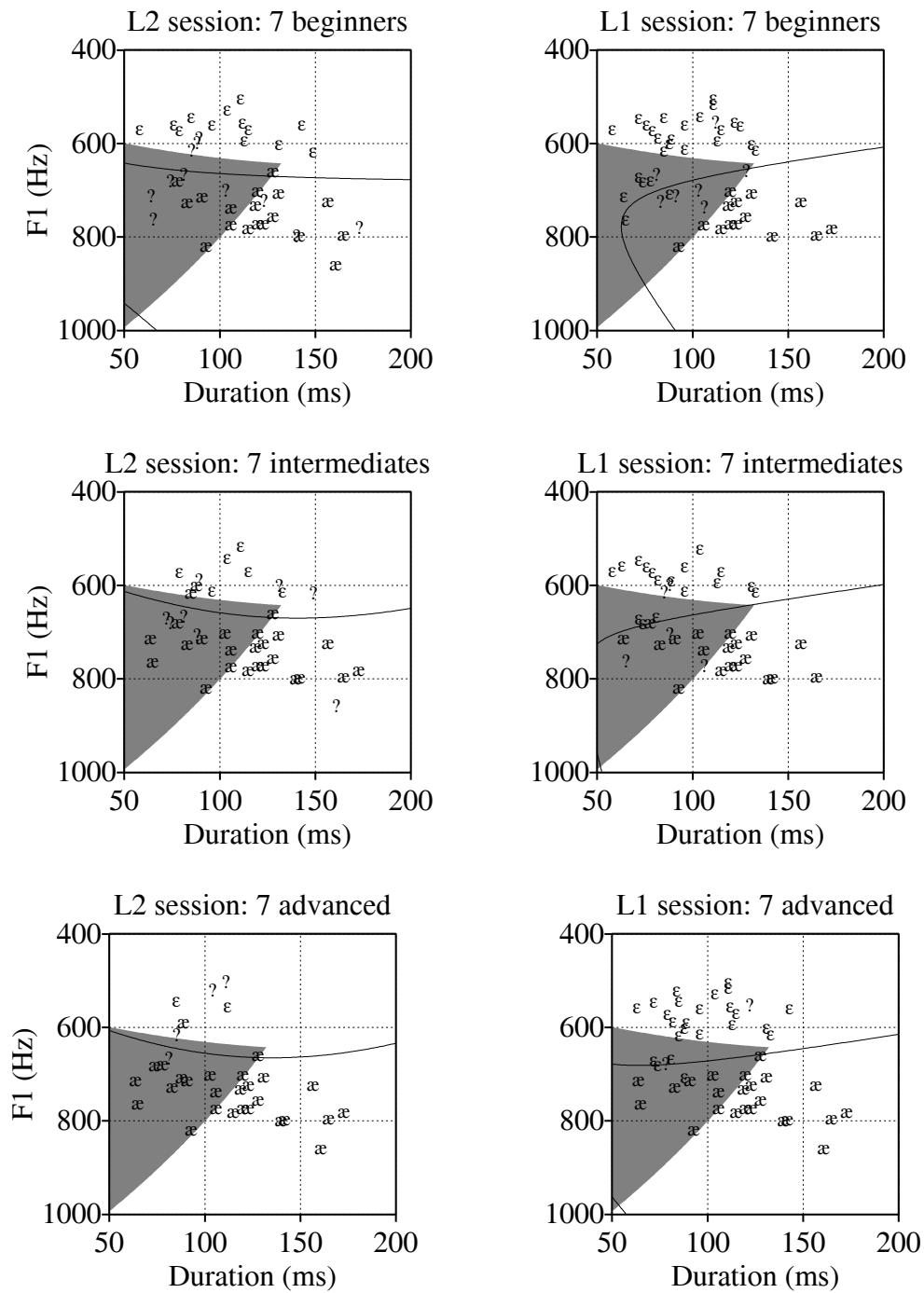


Fig. 15. L2 (left) and L1 (right) categorization for the three groups of CE learners of CF.

When looking at plot on the right of the figure, we can observe that none of the learner groups use duration differences when categorizing the L2 vowels, as shown by the horizontal shape of their category boundaries. This means that these learners needed just a little exposure to CF in order to acquire the optimal cue reliance, as can be inferred from the comparison of the first square on the left of Figure 15 with the CF boundary shown in Figure 6, above in §2.1. Crucially, we can also observe a developmental adjustment in the location of the F1 boundary; The beginning learners incorrectly categorize F1 values in the grey region as CF /ε/, whereas the intermediate and advanced learners correctly categorize almost

all tokens in the same region. Thus, taking the perception of the three groups together, it can be observed that learners categorize more and more L2 /æ/ vowels from the grey region as the optimal one, i.e. /æ/, as a function of their exposure level. This is visualized in Figure 16 below.¹² This means that CF /æ/ tokens produced with low F1 values and short durations that are categorized as /ɛ/ in monolingual CE perception are now being categorized as /æ/ in L2 CF. A ranked correlation test performed on the number of /æ/ responses for tokens that fall in the grey region and the learner's exposure level yielded a significant result (one-tailed Kendall's tau-b = 0.45, $N = 21$, $p = 0.004$, i.e. p from zero = 0.23%), which means that the observed development is statistically reliable.

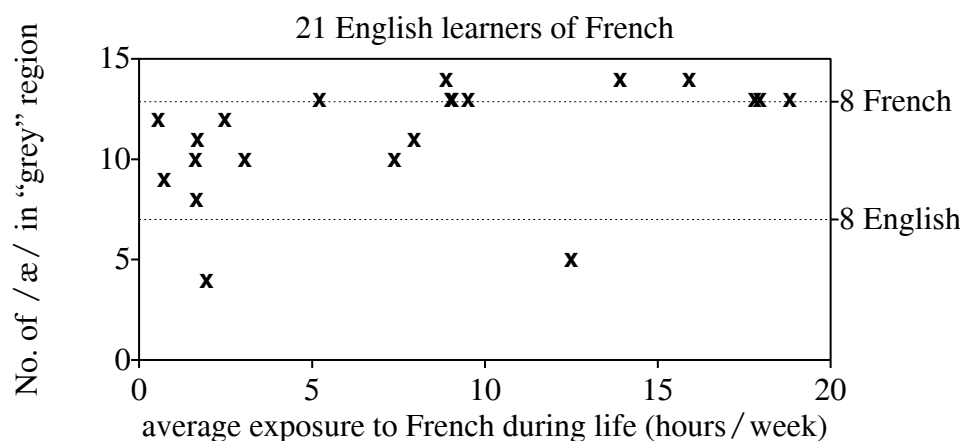


Fig. 16: Categorization of the 15 CF /æ/ tokens that fall in the grey region during the learners' French session (L2 session). Horizontal dotted lines: Average monolingual CF and CE perception (8 listeners per language).

3.5 L2LP ingredient 5: The L2 end state

The interrelation between the L1 and the L2 perception systems can constrain the L2 end state as well as the L1 perception after L2 development has occurred. Cook (2002), following Francis (1999), proposes that there are three logical possibilities for how the representations of two language systems interact in the mind of a second language learner. Figure 17 shows an adapted version of Cook's graph for the possibilities (2002: 11). In a *separate* systems view, L1 and L2 sound categories are thought to belong to autonomous systems. The *mixed* view advocates that L1 and L2 sound systems are, in fact, a single representational system. This perspective has, in turn, two possibilities, namely *merged* and *integrated* systems: Merged representations imply no language differentiation, whereas integrated representations refer assumes that the two languages are represented within the same system but are differentially tagged. Finally, in the *connected* view, L1 and L2 representations are mostly distinct but they may share some elements or properties.

¹² As can be seen in the figure, one of the subjects in the intermediate group exhibits an unexpectedly low number of /æ/ responses. This seems to suggest that there are individual differences in the speed in which learners can achieve native like performance in the learning of L2 similar sounds. An longitudinal study is needed to investigate the further development not only of this single outlier but also that of the other subjects who seem to, at this point in their acquisition process, conform to the model's predicted developmental path.

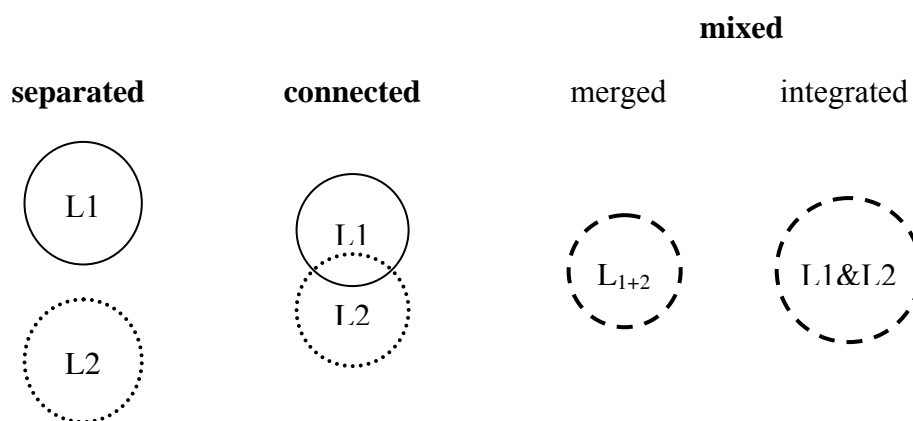


Fig. 17. Possible cognitive status of sound categories and perception processes in L2 learners (adapted from Cook 2002).

The L2LP model advances the separate perception grammars hypothesis which states that L2 learners and bilinguals have separate systems for perceiving their two languages, i.e., the left most option in the figure. In contrast, Flege’s (1995) Speech Learning Model (SLM) suggests that the perception and representation of L1 and L2 sounds is handled within a common L1-L2 phonological space (p. 239) in which sounds from the two languages coexist, as in the integrated system view. These different views yield two different predictions for the L2 end state. That is, the L2LP predicts that an L2 learner can attain optimal L1 and L2 perception because they are handled by two different systems, whereas the SLM predicts that any L2 development will inevitably affect the L1 because L1 and L2 development occur within a common space that gets adjusted by L2 or L1 changes.

A key construct that can be used to evaluate these contradictory predictions may be Grosjean’s (2001) hypothesis of the bilingual’s *language modes* hypothesis which is defined as “the state of activation of the bilingual’s languages and language processing mechanisms at a given point in time” (p. 2). According to this hypothesis the bilingual’s languages can be activated selectively or in parallel depending on a number of linguistic and extra-linguistic variables, such as the language of the experimenter, the task, the stimuli, the instructions, etc., which Grosjean defines as a the state of continuum between a monolingual mode and a bilingual mode. The L2LP interpretation of this hypothesis presupposes that L2 learners and bilinguals exhibit different language modes as a result of the activation of separate perception grammars during online perception. Following this interpretation, an advanced L2 learner is predicted to have an optimal perception in the monolingual setting of each language.

Thus, in a SIMILAR scenario, learners are predicted to be able to adjust their L2 category boundaries without affecting their already optimal L1 boundaries, as shown in Figure 18. Importantly, these language setting effects are particularly relevant in the SIMILAR learning scenario because here we can test whether learners use the same number of categories while demonstrating different perceptual behaviour in the monolingual L1 and L2 conditions.

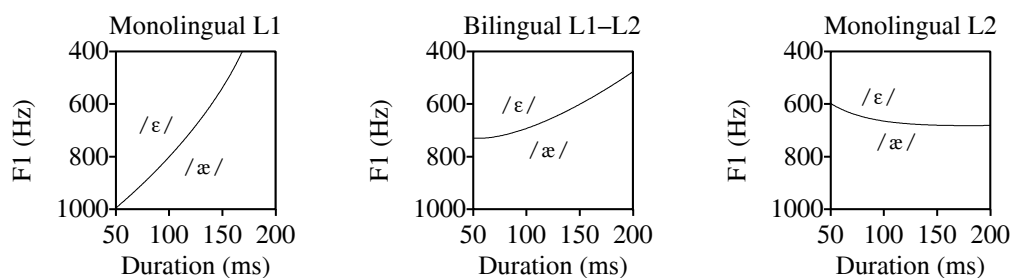


Fig. 18: Predicted three different types of perceptual behaviour in advanced Canadian English learners of Canadian French.

With respect to the available evidence in favour or against the hypothesis of separate grammars, Caramazza et al. (1973), Elman et al. (1977), and Flege & Eefting (1987) found that bilinguals and L2 learners exhibit perceptual category boundaries that are at an intermediate location when compared to the monolingual L1 and L2 boundaries. This finding can have two possible interpretations. It can be said that the bilingual and L2 listeners possessed a single grammar or it can be said that they possessed two grammars that were both activated during the perception experiments. If the latter is the correct interpretation, it should be possible to gather L1-like and L2-like category boundaries from the same learners when they are conditioned to use only one of their languages. If this can be done, it would mean that the intermediate L1-L2 boundaries do not represent a property of sound perception in bilinguals and L2 learners, but they are the result of performance.

Thus, according to the L2LP model, the finding of intermediate perception boundaries results from the parallel activation of two grammars during online perception, and does not, therefore, confirm the existence of a single grammar. The L2LP model proposes that L1 and L2 are handled by two different grammars with the same constraints but different rankings. At any time, either of these two grammars can be activated for auditory input, depending on the linguistic and paralinguistic evidence for the activation of one language system and the inhibition of another. For ambiguous auditory events, i.e. tokens which belong to the distributions of both languages, the output of both grammars will be equally likely to become the chosen candidate. This is because in 50% of the cases the winner of one grammar will be chosen, and in 50% of the cases the winner of the other grammar will be chosen. Consequently, the category boundary between vowels which have similar distributions in the L1 and the L2 can exhibit intermediate properties between monolingual categorization, if the language of the incoming token is not clear or if the listener is in a fully bilingual mode. The result of the activation of two perception grammars, due either to ambiguous tokens or to an ambiguous language setting, is depicted in the middle plot of Figure 18.

Escudero (in preparation) presents data which support the hypothesis of separate perception grammars for bilinguals and L2 learners. In this study, the author conditioned CE learners of CF to use only one of their languages in two different testing sessions, namely a monolingual L1 session and monolingual L2 session. In each session, the learners listened to the same CF tokens of /æ/ and /ε/ embedded in the language condition of the session. For instance, in their French session, they were tested by a French experimenter, were addressed in French only, and were told that the stimuli they would hear were French. In addition, prior

to the target perceptual experiment, the listeners were presented with a French passage and had to answer five general comprehension questions, all in French. This first task lasted for approximately 10 minutes and was used to enhance the use of French only. In the English session, the learners answered language background questions posed by a monolingual English-speaking experimenter, an activity that lasted 10 minutes and was performed prior to the English perception experiment.

Escudero's results show that the perception of the same CF tokens presented in the monolingual L1 condition turned out to be very different from the perception found in the monolingual L2 condition, as can be seen when comparing the squares on the right of Figure 15, above in §3.4, to the squares on the left. Figure 19 shows the learners' categorization of the CF /æ/ tokens that were produced in the grey region when listening to them as English, as opposed to the results of listening to them as French (Figure 16 above).

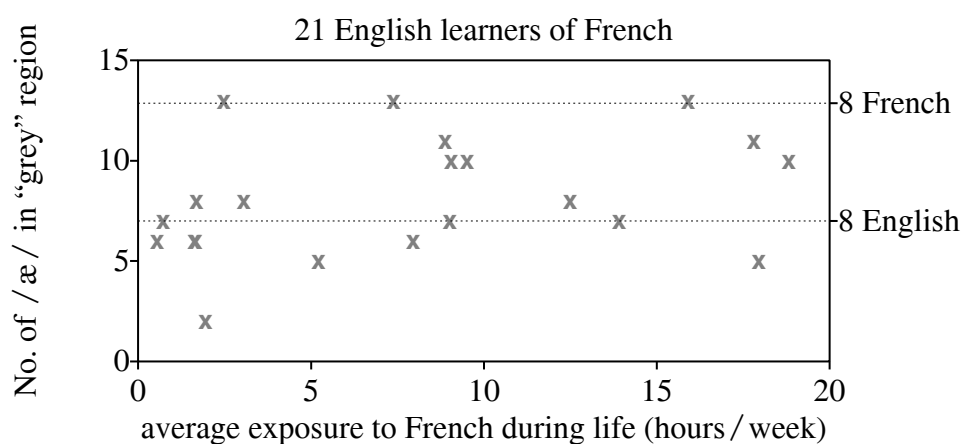


Fig. 19: Categorization of the 15 CF /æ/ tokens that fall on the grey region category during the learners' English session (L1 session). Horizontal dotted lines: Average monolingual CF and CE perception (8 listeners per language).

When comparing the results of Figures 16 and 19, it is easily observable that the learners perceive the same tokens differently depending on the language condition. Thus, the same ranked correlation performed for Figure 16 yields no effect of experience level when performed on the data shown in Figure 19. In addition, a paired-samples test conducted on the number of /æ/ responses in the results of the two language conditions confirms that the difference between listening to the same 15 CF tokens as French or as English is highly significant ($t = 4.51, N = 21, p < 0.0001$).

Thus, these findings suggest that L2 learners can achieve native-like L2 competence while maintaining their L1 perception in its original state. This means that L2 speakers have different ways of perceiving their two languages, L1 and L2, suggesting that they have two different perception grammars for them. Consequently, it can be said that the L2LP model's hypothesis that L2 speakers have two separate grammars is borne out and that the intermediate boundaries found in previous studies result from the activation of two separate grammars in a bilingual setting.

4 Conclusions

In this paper, I argued that speech perception is a linguistic phenomenon that should be brought into the domain of phonological modelling. This claim is based on the phonetic and psycholinguistic evidence which shows that adult speech perception is shaped by experience with a specific language, making it exclusively appropriate for that specific language. Further, I proposed a number of phonological, phonetic, and psycholinguistic criteria that a model should incorporate in order to arrive at a highly comprehensive and explanatory model for sound perception. Subsequently, I demonstrated that the LP model complies with these criteria. Importantly, the L1 acquisition extension of the LP model turned out to successfully lay out the mechanisms involved in learning to perceive the sounds of a language. With respect to the main objective of this paper, viz., explaining L2 sound perception, the L2LP model was shown to successfully describe, explain, and predict the learning of SIMILAR L2 sounds. This is summarized in Table 3.

L2LP ingredients	Predictions for SIMILAR	Finding
Optimal L1 & L2	CE and CF monolingual listeners will exhibit optimal L1 perception	Borne out
Initial state	Beginning CE learners will be equal to monolingual Spanish and CE listeners	Partially borne out
Learning task	Boundary shift for CE learners	Borne out
Development	Lexicon-driven learning	Indirectly borne out
End state	CE learners will attain optimal L2 perception and will maintain their optimal L1 perception	Borne out

Table 3: The five L2LP predictions for a SIMILAR L2 sound perception scenario and the evidence to support them.

Thus, I have shown that the L2LP model makes specific and explicit predictions for the perception of SIMILAR L2 sounds. First, the model predicts that listeners are optimal perceivers of their native language, a prediction that was borne out in the perception of /æ/ and /ɛ/ by monolingual CE and CF listeners. Second, it predicts that beginning L2 learners start with a copy of their L1 perception grammar and L1 perceived categories, a prediction that was confirmed by the perception of CE learners of CF because these learners made use of two L1 categories to perceive two L2 categories. Third, it was predicted that the learner would adjust her L1 perception to become an optimal L2 listener. It was found that the learners' perceptual boundaries of F1 and duration indeed gradually shifted in the direction of the boundaries for L2. Crucially, the model provides a formalization of the learning mechanism that leads to perceptual boundary shifting.

Finally, the L2LP hypothesizes that both L1 and L2 can be optimal because they are handled by two separate grammars. The data presented showed that CE learners of CF manifested a significant difference between their L1 and L2 perception of the same vowel tokens, thus confirming this prediction of the L2LP model. However, an even more rigorous procedure (especially with respect to the nature of the stimuli presented) is required to show whether the L1 perception of L2 learners remains monolingual-like and whether the difference between L1 and L2 perception increases with L2 exposure.

In sum, it can be concluded that the L2LP phonological proposal for L2 sound perception currently provides the most comprehensive description, explanation, and prediction of L2 sound perception. In this paper, it has been shown that the model can successfully handle the acquisition of L2 sounds that are phonemically equivalent but phonetically different from L1 sounds, which are also called SIMILAR L2 sounds. For the way the model handles other L2 sound learning scenarios such as NEW and SUBSET L2 sounds the interested reader is referred to Escudero and Boersma (2004) and Escudero (2005).

Acknowledgements

I would like to thank Silke Hamann and an anonymous reviewer for their extremely helpful comments and suggestions for improvements on the manuscript. The remaining possible mistakes are obviously my own.

References

- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, 171–203. Timonium, MD: York Press.
- Boersma, P. (1998). *Functional phonology*. Doctoral Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
- Boersma, P., P. Escudero & R. Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013-1016. Barcelona: Causal Productions.
- Boersma, P. & B. Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.
- Broselow, E. (2004). Language contact phonology: richness of the stimulus, poverty of the base. In K. Moulton & M. Wolf (eds.), *Proceedings of the North Eastern Linguistics Society (NELS) 34*. Amherst, MA: GLSA.
- Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136-193.
- Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell.
- Bürki-Cohen, J., J. L. Miller & P. D. Eimas (2001). Perceiving non-native speech. *Language and Speech* 44, 149-169.
- Caramazza, A., G. Yeni-Komshian, E. Zurif & E. Carbone (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 5, 421–428.
- Cook, V. J. (2002). Background to the L2 user. In V. J. Cook (ed.), *Portraits of the L2 User*, 1-28. Clevedon: Multilingual Matters.
- Dupoux, E., C. Pallier, K. Kakehi & J. Mehler (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 5, 491-505.

- Elman, J., R. Diehl & S. Buchwald (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* 62, 971-974.
- Escudero P. (2001). The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels. In A. H.-J. Do, L. Dominguez & A. Johansen (eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development*, 50-261. Somerville, MA: Cascadilla Press.
- Escudero P. (2005). *Linguistic Perception and second language acquisition*. Doctoral Dissertation, Utrecht University.
- Escudero, P. (in preparation). Evidence for gradual L2 re-categorization: Boundary shifts in the L2 perception of Canadian French /æ/.
- Escudero, P. & P. Boersma (2001/2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In S. Arunachalam, E. Kaiser & A. Williams (eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8, 71-85.
- Escudero, P. & P. Boersma (2004a). L2 re-categorization of an 'old' phonological contrast. Poster presented at the 9th Laboratory Phonology Conference, University of Illinois at Urbana-Champaign.
- Escudero, P. & P. Boersma (2004b). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26, 551-585.
- Escudero, P. & L. Polka (2003). A cross-language study of vowel categorization and vowel acoustics. In M. J. Sole, D. Recansens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 861-864. Barcelona: Causal Productions.
- Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, 233-277. Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production*. Berlin: Mouton de Gruyter.
- Flege, J. E. & W. Eefting (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English. *Speech Communication* 6, 185-202.
- Gerrits, E. (2001). *The categorisation of speech sounds by adults and children*. Doctoral Dissertation, University of Utrecht.
- Gottfried, T. L & P. S. Beddor (1988). Perception of temporal and spectral information in French vowels. *Language and Speech* 31, 57-75.
- Grosjean, F. (2000). The bilingual's language modes. In J. Nicol (ed.), *One Mind, Two Languages: Bilingual Language Processing*, 1-22. Oxford: Blackwell.
- Hillenbrand, J. M., M. J. Clark & R. A. Houde (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America* 108, 3013-3022.
- Hume, E. & K. Johnson (2001b). A model of the interplay of speech perception and phonology. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 3-26. New York: Academic Press.
- Hyman, L. M. (2001). The limits of phonetic determinism in phonology: *NC revisited. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 141-186. New York: Academic Press.
- Jacquemot, C., P. Pallier, D. LeBihan, S. Dehaene & E. Dupoux (2003). Phonological grammar shapes the auditory cortex: a Functional Magnetic Resonance Imaging study. *Journal of Neuroscience* 23, 9541-9546.
- Jones, C. (2003). *Development of phonological categories in children's perception of final voicing*. Doctoral Dissertation. University of Massachusetts at Amherst.
- Jusczyk, P. W., A. Cutler and N. J. Redanz (1993). Infants' preference for the predominant stress patterns of English words. *Child Development* 64, 675-687.
- Kenstowicz, M. (2001). The role of perception in loanword phonology. *Studies in African Linguistics* 32, 95-112.
- Kirchner, R. (1999). Review of P. Boersma (1998) *Functional Phonology*. *GLOT International* 4 (9/10), 13-14.
- Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnetic effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93-107.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA* 97, 11850-11857.

- Major, R. C. (2001). *Foreign Accent*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Maye, J., J. F. Werker & L. A. Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.
- McQueen, J. M. (2005). Speech perception. In K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition*, 255-275. London: Sage Publications.
- Miller, J. L. & E. R. Dexter (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 14, 369-378.
- Miller, J. L. & F. Grosjean (1997). Dialect effects in vowel perception: the role of temporal information in French. *Language and Speech* 40, 277-288.
- Miyawaki, K., W. Strange, R. R. Verbrugge, A. M. Liberman, J. J. Jenkins & O. Fujimura (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18, 331-340.
- Nittrouer, S. & M. E. Miller (1997). Developmental weighting shifts for noise components of fricative-vowel syllables. *Journal of the Acoustical Society of America* 102, 572-580.
- Pitt, M. A. & J. M. McQueen (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language* 39, 347-370.
- Polka, L. & J. F. Werker (1994). Developmental changes in the perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20, 421-435.
- Prince, A. & P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science.
- Schachter, D. L. & B. A. Church (1992). Auditory priming: implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18, 521-533.
- Steriade, D. (2001). Directional asymmetries in place assimilation: a perceptual account.. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 219-250. New York: Academic Press.
- Strange, W. (1995). Cross-language study of speech perception: a historical review. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 3-45. Timonium, MD: York Press.
- Tesar, B. & P. Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Werker, J. F. & J. S. Logan (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37, 35-44.
- Werker, J. F. & R. C. Tees (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development* 7, 49-63.