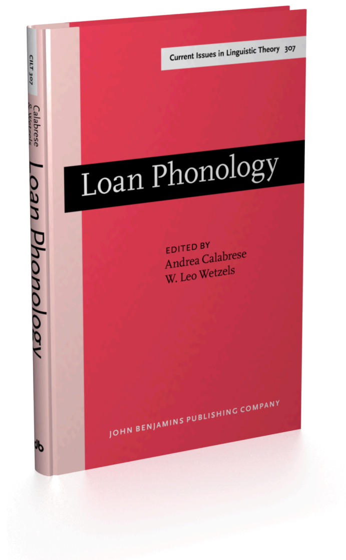
Current Issues in Linguistic Theory 307

Perception, production and acoustic inputs in loanword phonology

Andrea Calabrese

https://doi.org/10.1075/cilt.307.03cal



Pages 59–114 of
Loan Phonology
Edited by Andrea Calabrese and W. Leo Wetzels
[**Current Issues in Linguistic Theory**, 307] 2009. vii, 273 pp.

John Benjamins Publishing Company

Perception, production and acoustic inputs in loanword phonology*

Andrea Calabrese

1. Introduction

This work started as a purely linguistic investigation of the phonological adaptations found in loanwords. However, the importance that speech perception plays in accounting for them soon became apparent, and a number of interesting observations resultantly came to light. Accordingly, this paper has become a study of both loanword phonology and speech perception. Central to this topic is the issue of how one perceives and learns unfamiliar sound configurations, or better, words containing unfamiliar sound configurations, and how these sound configurations are adjusted during this process.

As we will see, loanwords are generated by bilinguals when they take words from one of the languages they know and use them in another of the languages they know. In this case the adjustments that the loanwords undergo occur during speech production. However, there is still another situation in which loanwords are produced: monolinguals may learn new words from a language they don't know, or know poorly, to fill a lexical gap in their language. In the present study I will focus on loanwords generated in this way. The issues of speech perception, and of word learning, are fundamental in this latter case, and will therefore be central in the analysis in this paper.

Generally speaking, the goal of speech perception is the determination of the meaning of an utterance that generated a given acoustic input. This is achieved by identifying the words present in the utterance, and establishing their syntactic organization. When we want to learn foreign words, or even new words in our own language, however, the main goal of speech perception is the identification of their phonological shape so that we can properly memorize them. This involves

*I would like to thank Jonathan Bobaljik, Sylvain Bromberger, Morris Halle, Michael Kenstowicz, Andrew Nevins, Donca Steriade, Leo Wetzels, and audiences at the Universities of Siena and Sassari for criticisms, comments and suggestions.

constructing—by means of an inferential computation— a mental representation of the words in terms of articulatory features, including fully specified syllabic and prosodic structures.

I will argue that identifying, or perhaps more aptly, "understanding", a sound, or a combination of sounds, means identifying the "instructions", i.e. the feature configurations,[1] characterizing its articulation in the production component. When a learner is faced with a new language, he must deal with sounds that are not included in the inventory of his language. The featural configurations characterizing these segments or combinations of segments are therefore absent in the production system, and the foreign sounds cannot be articulated.[2] It follows that

1. See later in this section on features as "instructions".

2. In Calabrese (1995, 2005) this means that there are active marking statements forbidding these feature configurations/combinations.

In this paper I will try to avoid using any particular formal theory of phonology. However, my own way of seeing phonology (see Calabrese (2005)) will surely influence some of my theoretical choices. To make those theoretical choices more clear, I will give a brief synopsis of the theoretical beliefs relevant to this work in this footnote.

The phonological system of a language is a historically determined complex set of output phonological representations derived from mnemonic representations by phonological operations. The input and output representations of the derivation must be such that they are able to interface properly with the relevant physical/mental component. Therefore, output representations must be able to be properly articulated by the motor system and properly perceived by the sensory system. Input representations must be such that they can be encoded in long-term representations in the memory. The proper interface properties of output representations, i.e. their ability to be pronounced and perceived, are determined by the constraints and rules contained in the markedness module. These constraints and rules trigger operations that convert illicit configurations into licit configurations that can be interpreted by the sensory-motor system.

The markedness module includes universal negative constraints such as the *prohibitions* and *marking statements*. Prohibitions identify configurations that are never possible for articulatory and/or acoustic/perceptual reasons. An example is *[+high, +low] which is necessarily articulatorily impossible insofar as the tongue body cannot be raised and lowered simultaneously.

Marking statements identify phonologically complex configurations that may be found in some but not all phonological inventories. An example is *[−back, +round] which "marks" the feature configuration [−back, +round] as phonologically complex. Another is *Complex Onsets which "marks" complex onsets as difficult. The reasons for their complexity or difficulty are due to independent properties of the sensory motor system that are reflected in the grammar through these constraints (see below). Marking statements may be active or deactivated. If a marking statement is active in a language, the complexity of this configuration is not accepted in this language, and segments containing this configuration are absent from the language.

the learner cannot "understand" them when they are heard. He lacks a mental representation of them. Through auditory exposure and motor training, the learner can learn to produce them, and therefore "understand" them in the perception process. However, such learning is difficult and time consuming. The other way is to adjust their featural representations to make them familiar, and therefore the learner can "understand" them in perceptual mental representations. These adjustments are implemented through the same repair operations used in the production system. The perceptual mental representation that is so obtained is then stored in long term memory, and becomes the adjusted underlying representation for the non-native sound configuration.

I assume that this process is common to all experiences of non-native sounds, in particular learning a new language or borrowing words with foreign sounds or sound combinations from a different language. In the latter case, given that there is no need to preserve the phonological and morphological shape of the foreign word, as in language acquisition, the foreign word can fully undergo adjustments that can be both phonological and morphological.

If perception involves interpretation and computation, as discussed above, it loses its primary function of tracking external reality, the environment; it becomes detached from reality and prone to illusions. Although illusion-like, interpretative failures may occur, as discussed in later sections, I assume that listeners always have a direct access to the acoustic signal. I will further assume that a representation of the acoustic signal is stored in a short-term acoustic memory ("echoic

---

If a marking statement is active in a language, the configuration marked by this statement is not accepted in this language. This configuration is *illicit*. Illicit configurations are fixed up by set of repairs provided by UG.

In this text when a segment or another structure cannot be articulated because it is "foreign" this implies the existence of marking statements characterizing the segment or structure in question as illicit; therefore it requires repair.

Marking statements and prohibitions belong to the grammar; they are grammatical statements about phonological representations. However, they are also interface conditions, i.e. the means through which the linguistic computational system is able to interpret and read the properties of the sensory-motor system. The markedness constraints represent the sensory-motor system in the linguistic computational system.

In particular, active marking statements indicate the absence, or unavailability, of computational programs converting phonological representations into articulatory ones. When a marking statement becomes active, the targeted phonological configurations cannot be transformed into articulatory commands; the repair procedures that occur in this case must then refer to the manipulations of phonological configurations that make this transformation possible (see Calabrese (2005) for more discussion of this model).

memory" see Neisser (1967)). I will propose that there is also an echoic long term memory, and argue that the acoustic representations preserved in echoic memory tie perception to external reality.

We will see evidence for two perceptual systems. One is dedicated to picking up environmental information, bottom-up; the other one is the reverse, functioning top-down, and is dedicated to analysis, identification, and recognition. The first system is tuned to the environment. In the case of speech, it collects and stores the acoustic signal—echoic memory is part of it. This system implements the acoustic analysis of the signal extracting its invariant spectral properties. It also discriminates new, unfamiliar sounds and sound strings from familiar, previously heard ones. What is new is sent for further analysis to the second perceptual system while it is temporarily stored for possible comparisons. The second system is then active in analyzing those new or unfamiliar configurations. The production component plays a basic role in it insofar as this system analyzes linguistic material by synthesizing it anew—it is an analysis-by-synthesis system (Halle & Stevens 1962). This system is fundamental in learning insofar as analysis of what is new is crucial for learning.

As in Calabrese (2005), here I assume a realistic approach to language (Bromberger & Halle 1992, 1997, 2000; Halle 2002), according to which. "phonology is about concrete mental events and states that occur in real time, real space, have causes, have effects, [and] are finite in number." (Bromberger & Halle 2000:21). If we look at production, this realistic view of language assumes that phonological theory investigates the system of knowledge that allows concrete occurrences of real time computational steps that convert mnemonic representations of utterances into articulatory representations. This knowledge involves representations and computations that have concrete spatio-temporal occurrences allowing for the production of concrete articulatory events which stem from the workings of an actual brain with all its limitations.

When we turn to perception, especially if we interpret the term "realism", naively, such an approach should lead us to focus on the concrete reality of the linguistic signal that is perceived. This reality is acoustic. Perception should then simply involve **bottom-up** processes that extract all the relevant perceptual information from the acoustic input as it comes in, without recourse, or with minimal recourse, to **top-down** processes involving independent linguistic knowledge. All of the information needed for the identification of the words and morphemes contained in an utterance should be present in the acoustic signal according to this view.

Evidence shows that this concept cannot be maintained. Problems with this idea are brought to light by Liberman (1957). One very striking finding in his research was that, due to coarticulation, acoustic cues for consonants especially are highly context sensitive. For example, take the syllables /di/ and /du/. The information critical to the identification of these syllables is the transition of the second formant. However, that transition is high in frequency and rising in /di/, but low and falling in /du/. In the context of the rest of each syllable, the consonants sound alike to listeners. Separated from context, they sound different, and they sound the way they should sound: two "chirps," one high in pitch and one lower. Acoustically, they do not have a plausible common denominator, an invariant property, despite the fact that they are perceived as the same sound. Liberman recognized that, despite the context sensitivity of the acoustic signals for /di/ and /du/, naturally produced syllables do have one thing in common: they are produced in the same way. Therefore, articulatorily, the consonants have a single common denominator. Both syllables are produced by using the tongue tip to make a constriction in the alveolar region. Listeners' percepts are based on the articulatory reality of sounds.

Facts such this show that the identification of sounds in acoustic inputs needs information that is not immanently contained in these inputs, and that therefore cannot be simply extracted **bottom-up** from them, rather must be computed—**top-down**—through processes that can access the production system.[3] Notice that this does not contradict the realistic approach to language proposed above which assumes that linguistics deals with concrete mental events and states. In this view, both bottom-up and top-down processes involve concrete mental events and states that are an organic part of the perceptual experience.

Top-down processes in perception are also needed for other reasons. We know that listeners may "restore" missing phonetic segments in words (Samuel 1981; Warren 1970), and talkers shadowing someone else's speech may "fluently restore" mispronounced words to their correct forms (see Marslen-Wilson & Welsh 1978). This ability to restore missing phonemes or correct erroneous ones can be explained only if we assume top-down processes that access information in the lexical entries. Even more significant departures of perceptual experience from the stimulus may be observed in some mishearings (for example "popping really slow" heard as "prodigal son" (Browman 1980; Fowler 1986)) or "mow his own lawn" heard as "blow his own horn" (Garnes & Bond 1980; Fowler 1986). As for mishearings, Garnes and Bond (1980) argue that "active hypothesizing on the part of the listener concerning the intended message is certainly part of the

3. The computational module that performs these operation was called the *phonetic module* by Liberman and Mattingly (1985).

speech perception process. No other explanation is possible for misperceptions which quite radically restructure the message … ” (p. 238).

Access to the production system, to lexical information, and the ability to reconstruct, or even construct phrases and sentences in the percept indicate that perception clearly involves top-down processes. In this work, as mentioned previously, I propose that a top-down system plays a fundamental role in speech perception. The speech production component is part of this top-down system. This is the system that implements the analysis-by-synthesis of linguistic inputs.

One could propose, as in the motor theory of perception (Liberman & Mattingly 1985; Liberman & Whalen 2000), that perception is essentially a top-down process of construction and interpretation where bottom-up processes accessing the acoustic input have a minimal role. In this article I will also argue against this hypothesis, and point out a fundamental problem faced by top-down perception: if perception is essentially analysis through production, we should expect that sounds that cannot be articulated in production, e.g. foreign sounds, should also be unperceivable. Therefore, a learner could never be able to access them and learn them. This is contrary to the common human experience of foreign sounds. A learner can hear a foreign sound even though he cannot articulate and recognize them, and will try to learn them—"apprehend" them, as proposed later—by constructing articulatory representations that approximate their acoustic reality. The acoustic reality of the speech input must also be accessible in perception. To account for this fact, I will propose that a fundamental role in perception is played by echoic memory where acoustic images, that is, acoustic representations of inputs, are stored. As proposed above, echoic memory is part of the bottom-up component of perception where a preliminary acoustic analysis of the speech input is implemented. Thus following a model such as Trace (McClelland &Elman 1986; see also Klatt 1980), but in a strictly computational formulation, I will argue that perception must contain both a bottom-up and a top-down component that run in parallel and interact with each other.

Finally, a fundamental assumption of the present paper which must be made explicit before concluding the introduction is that in long term memory morphemes and words are represented as sequences of discrete segments each of which is characterized by a bundle of distinctive features. There is overwhelming phonological evidence that this is the correct interpretation, though I will not discuss this evidence here, but refer to Kenstowicz (1994), Halle (2002) and others for the compelling phonological arguments supporting this view. It is also important to highlight that acoustic evidence also supports such a view. Acoustic studies (Stevens 1972, 1989, 2002) of sounds produced by various manipulations of the vocal tract show that certain distinctive and stable acoustic patterns occur when the vocal tract is in particular configurations or performs particular maneuvers—these configurations or maneuvers correspond to distinctive features. As Stevens (2002) points out, these combinations of acoustic and articulatory patterns are based on the physics of sound generation in the vocal tract, including theories of coupled resonators, the influence of vocal-tract walls on sound generation, and discontinuities or stabilities in the behavior of sound sources. Evidence for features also comes from quantal aspects of auditory responses to sound, such as responses to acoustic discontinuities and to closely-spaced spectral prominences (Chistovich & Lublinskaya 1979; Delgutte & Kiang 1984; Stevens 2002).

Distinctive features have a dual function. First, they serve as mnemonic devices that distinguish one phoneme from another in speakers' memories, a function that is fundamental during speech perception. Each feature also serves as an instruction for a specific action of one of the movable parts of the vocal tract, a function that is fundamental during speech production (cf. Halle 2002).

The phonetic substratum for each feature establishes a link between a specific articulatory action and an acoustic and perceptual consequence of this action. As proposed by Liberman and Mattingly (1985, 1989), Halle (2002), Halle and Stevens (1991), the computational system that makes them capable of acquiring command of one or more languages includes a module, which I will assume is part of the top-down perceptual component, that selects specific actions of the articulators and links them to selected aspects of their acoustic consequences (Halle &Stevens 1991). These correlations between articulatory activity and acoustic signal are controlled by the distinctive features. For example, the forward and backward placement of the tongue body is correlated with specific differences in the frequency of the second formant – this correlation is controlled by the feature [back]. Other examples include the correlations of the different placements of the tongue blade, be it before or behind the alveolar ridge, with the differences in the acoustic spectrum between hissing and hushing sounds, a difference controlled by the feature [anterior]. Similar relations between articulatory activity and acoustic signal are provided for each of the roughly nineteen features that comprise the universal set of phonetic features (Halle 1992; Halle & Stevens 1991).

In the next section, I will briefly review the most recent theoretical models in loanword phonology. After this brief review, I will discuss the perception model proposed here and demonstrate that this model offers the most adequate account for the adaptations of the foreign sounds found in loanwords.

2. Loanwords

I begin by considering the nature of loanwords. First of all, one can distinguish two types of loanwords: *integrated loanwords* and *on-line adaptations* (Peperkamp

2002). I*ntegrated loanwords* are words that have entered the lexicon of the borrowing language. Monolingual speakers who use these loanwords never hear their source forms, and so the phonological analysis of the modifications these words have undergone when entering the borrowing language has no direct psychological reality. Rather, it receives a diachronic interpretation, in that it accounts for the adaptations applied by those speakers who originally introduced the loans. The *on-line adaptations* are foreign words that are borrowed 'here-and-now' (see, for instance, Shinohara (1997, 2000) and Kenstowicz & Sohn (2001)). In this paper, following Peperkamp, I treat integrated loanwords and on-line adaptations on a par, assuming that the former reflect on-line adaptations by those speakers who once introduced these words.[4]

Consider now the conditions under which linguistic borrowing occurs, i.e. the conditions that lead to the formation of loanwords. Assume two languages L1 and L2. L1 is the borrowing language and L2 the loaning language. Borrowing occurs when a speaker of L1 "borrows" a word of L2 to fill a lexical gap in L1. The reasons for this lexical gap can be many: lexical or cultural innovation may introduce objects or actions that do not have a name in L1; certain words may be felt as non-prestigious; certain words may simply be unknown, or just forgotten; new words may be created for playing, etc.

In any case, there are two possible scenarios in which borrowing occurs:

I. A speaker is bilingual in L1 and L2. A lexical gap in L1 is filled in by taking a word from L2. The speaker retrieves the underlying representation of this word from his L2 mental dictionary (the long-term memory storage for L2

---

4. In the case of the integrated loanwords, we also need to distinguish loanwords that underwent morphological nativization like the Italian loanwords (ia) from those that did not (b):

(i)	a.	bistecca (bistecc+a)	from	English *beef-steak*
		birra (birr+a)		German *bier*
		giulebbe (giulebb+e)		Arabic *gandulāb*
	b.	sport		
		jeep [dʒip]		
		killer		

The morphologically nativized loanwords in (a) are characterized by the addition of affixes and by other morphological and phonological changes characteristic of the native grammar. Often it is difficult to distinguish these words from those that etymologically belong to the native lexicon. These loanwords can be treated with the other integrated loanwords, but they also require an analysis of the processes of morphological nativization that applied to them. I will not deal with this type of processes here (see Repetti 2003, 2006, This volume, for an analysis of changes of this type in the English loans to Italian.)

lexical items) and generates its surface representation while speaking in L1. If the surface representation of the word is generated by using the phonological, or more generally the grammatical, system of L2, the word is pronounced as in L2. There are no adjustments or adaptations. However, if the surface representation of the word is generated by using the phonological, or more generally the grammatical, system of L1, the word undergoes adaptations and adjustments. It is nativized according to the L1 grammar.

Examples of this type of borrowing can be found in utterance below, directions that were given to me in the Boston North End by an Italian American shopkeeper when I asked him where I could park my car to get to his store, a notoriously difficult enterprise. We had known each other for a long time. His response to my question, which I asked in Italian, appears below:

(1) *Gira al corno di quella stritta. Poi prendi la seconda stritta a destra e vai stretto per due blocchi. Puoi parcare il carro proprio li.*

(2) Lexical Borrowings from English into Italian

English	**Loanword**	**Italian counterpart**
corner	corno	angolo
street	stritta	strada
straight	stretto	diritto
blocks	blocchi	——
park	parkare	parcheggiare
car	carro	macchina

He was born in the Campania region of Italy, and had come to the USA when he was a teenager. He spoke Italian pretty well, although in his regional accent.

In the example above, we can see that my friend replaced the appropriate Italian roots, which may have slipped his mind while speaking, with their English counterparts. The English roots were adjusted by modifying their phonology (in terms of changing vowel quality and the gemination of word-final obstruents) and by adding Italian suffixal morphology. These adjustments were obviously done on-line while he was producing his utterance.

II. The speaker of L1 does not know L2 well. He fills a lexical gap in L1 by learning the relevant word from a L2 speaker. Once the learned word will be uttered publicly or even silently, it becomes a loanword. Given that the speaker does not speak L2 well, the word will display adjustments and adaptations. There are two possible hypothesis to account for this. The first is that during perception and learning, the acoustic representations of the non-native segments is faithfully mapped into abstract featural representations (Jacobs & Gussenhoven (2000)). This featural representation is then modified during production. The second is that the modifications already occur during perception and learning. Below we will see evidence supporting the second hypothesis.

Therefore, adjustments and adaptations may occur during speech production, and during speech perception and analysis, and the situation appears to be identical in both cases. There is, however, an obvious difference in the inputs. In the first case, the input to the adaptations and adjustment is an abstract long-term memory representation, an underlying representation. In the second case, the input is the acoustic signal produced by the surface phonetic representation of the word. This difference, obviously, determines the final shape of the loanword. We will consider the consequences of this difference in the next section.[5]

3. The current major accounts of non-native phonological adaptations

There are two models of loanword phonology: one assumes that borrowing occurs only in scenario I; the other assumes that borrowing occurs only in scenario II.

3.1 The phonological repair model

This model proposes that nativization is brought about by the phonological processes characterizing speech production. Thus, LaCharité and Paradis (2005) (see also Paradis & Tremblay (this volume) and Jacobs & Gussenhoven (2000) as well as the analysis of loanwords in Calabrese (1988, 1995) and Connelly (1992)) attempt

5. It is important to note that during the process of adaptation, the adapter, exposed to a foreign sound from a language X, may also opt to acquire it as-is. Thus this sound may appear in the loanwords from language X. The same may occur for foreign syllabic and prosodic structures. For example, in Itelmen (J. Bobaljik (p.c.)), voiced stops occur only in loanwords. This type of innovation leads to situations in which the loanwords form X have a different phonology—different segments and possibly also other phonological and morphological phenomena derived from X—from that of the native words. They will thus form a special lexical stratum or co-phonology (Ito & Mester (1999). See Bobaljik (2006) for arguments in favor of this approach to loanword phonology.

Bobaljik (pc.) also referred me to Boberg (1991) who shows that for many speakers of American English there is a general phoneme "foreign A" [a:] which is used for foreign words like "pasta", "Mazda", "llama", "spa", "tobacco" regardless of source, but which are distinct from any other particular phoneme of (native) American English. As Bobaljik observes, in this case it is not surface phonetic similarity to the target that predicts the nativization pattern in the case of these words, just identification of them as belonging to a special [+foreign] class of words, and then regional and sociological factors.

I will not discuss cases of this type here.

to build production-based repairs into their nativization model. They assume that adapters start with underlying representations containing the non-native segments because the adapters are bilingual (LaCharité & Paradis 2005). Repairs to these non-native segments are implemented so as to avoid the production of marked or illicit segments or strings. A characteristic feature of these approaches is that speakers adapt loanwords by operating on a phonological/phonemic level that abstracts away from the details of allophonic and phonetic realization. The input to the adaptations is an abstract morphophonemic representation of the L2 word. For example, LaCharité and Paradis (2005) discuss the adaptation of English loans into French where the English lax high vowels /I/ and /U/ are mapped to French /i/ and /u/ instead of the acoustically closer /e/ and /o/. If loanwords are adapted in terms of operations on distinctive features then the configuration [+high, −ATR] of English /I/ and /U/ can be repaired into the configuration [+high, +ATR] of the French high vowels /i/ and /u/ regardless of the fact that French mid /e/ and /o/ are better acoustic matches in surface phonetic representations

Adaptations occur only during speech production in this model (cf. scenario I above).

3.2 Acoustic approximation model

According to this model, the adaptations we observe in loanwords are based on phonetic approximation/similarity. This model was first proposed by Hermann Paul in 1880. In his discussion of loanword phonology, he hypothesizes that a host speaker, upon encountering a foreign segment, matches this phonetic signal with the native segment with which it is most closely related. Paul implicitly assumes that this match involves a perceptual similarity judgment based on *Sprachgefühl*, the feeling of language: speakers adapt a non-native segment to one which they 'feel' most closely resembles the former acoustically.

New models of loanword phonology that acknowledge the importance of perception in determining similarity as the basis for the treatment of the loanwords (Silverman 1992; Yip 1993; Kenstowicz 2001, 2003 (see also Hsieh, Kenstowicz & Mou (this volume)) go back to this nineteenth century framework. According to them, the replacement operation between the non-native segment and the native one is strictly based on phonetic similarity between the outputs of the donor and recipient languages. Thus, according to Peperkamp and Dupoux (2003), the equivalences in loanword adaptation are based on a similarity that is defined as "acoustic proximity or proximity in the sense of fine-grained articulatory gestures."

In this model, the input to the adaptations is a surface phonetic representation of the L2 word and the similarity judgments producing the adaptations occur only during perception (cf. scenario II above)

3.3 Ito, Kang and Kenstowicz (2006)

Ito, Kang and Kenstowicz (2006) demonstrate that both models fail to account for the adaptation of Japanese vowels in Korean. The standard phonemic vowel inventories of the two languages are given in (3):

(3)	i	u	i	ɨ	u
	e	o	e, ö	ʌ	o
		a	ɛ		a
	Japanese		Korean		

The Japanese vowels are a subset of the Korean inventory. If we assume the acoustic/ perceptual model of nativization, given their different sizes, the vowels of the two systems might be expected to partition the articulatory- acoustic space differently. However, when we examine the loanword correspondences, we find that most vowels pick out exactly the phonologically matching Korean vowels (see (4)).

(4)	Japanese	Korean	
	beNtoo	pent*oo	'boxed lunch'
	azi	aci	'horse mackerel'
	hako	hak*o	'box'
	kagami	kakami	'mirror'
	sebiro	sepiro	'suit'
	teNpura	tenp*ura, temp*ura	'tempura'

There is one systematic exception. Japanese /u/ is adapted with the Korean central vowel /ɨ/ when it appears after the coronal sibilants [ts], [s], and [(d)z]. Elsewhere, it is adapted as Korean /u/ as shown in (5).

(5)		Japanese	Korean	
	a.	baNgumi	paŋkumi	'program'
		unagi	unaki	'eel'
		jurumi	jurumi	'relaxed'
		gaku	kak*u	'frame'
	b.	[ts]umi	s*ɨmi	'stack, pile'
		jaki[ts]uke	jak*is*ɨk*e	'glazing, baking'
		ko[ts]uzai	kos*ɨcai	'iron bar'
		susi	sɨsi	'sushi'
		suimono	sɨimono	'type of soup'
		mizuage	micɨake	'unloading catch of fish'
		kazunoko	kacɨnok*o	'herring roe'

This apparent exception is easily explained. Japanese /u/, frequently transcribed as the unrounded high back vowel [ɯ], is realized as centralized [ɨ] after [ts], [s], and [z] (Homma 1973:352–3; Fitzgerald 1996). This is an example where a nondistinctive variant in the source language coincides with a phoneme in the borrowing language (Iverson & Lee 2004).

Ito, Kang and Kenstowicz argue that although the Japanese /u/ lacks lip rounding, it is articulated with vertical lip compression (Vance 1987; Ladefoged & Maddieson 1996; Okada 1999). In other words, the Japanese high back vowel is produced with narrowing of the lip opening but without lip protrusion. The allophonic realization of the laryngeal fricative /h/ as a bilabial fricative [ɸ] before /u/ is indicative of this labial component. Here, to account for the these facts, I propose a modification of the feature [round], splitting it into two different features [labial] and [lip protrusion] where [+labial] is implemented by the narrowing of the lip opening. These two features are related by the marking statement in (6a) and the prohibition in (6b):[6]

(6) a. *[+labial, −lip protrusion]
b. **[−labial, +lip protrusion]

(6b) characterizes the configuration [−labial, +lip protrusion] as being always impossible. (6a), though, is characterized by high complexity and is therefore it is rarely deactivated (although there are some exceptions, like Swedish and Japanese). Thus, vowels with lip compression, like [+labial], are usually produced with lip protrusion, i.e. they are [+round]. However (6a) is deactivated in Swedish. Accordingly, a contrast between front rounded vowels with [+lip protrusion] and [−lip protrusion] can be observed in this language as in (7) (from Ladefoged & Maddieson (1996)):

(7)	+labial (lip compression)	+labial (lip compression)	−labial
	+lip protrusion	−lip protrusion	−lip protrusion
	ryːta	rʉːta	riːta
	‘roar’	‘window pane’	‘draw’

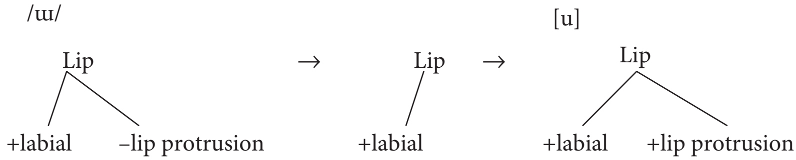
(6a) is deactivated in Japanese, but not in Korean. The feature specifications of the relevant vowels are given in (8):

(8)	a.	Japanese	i	ɨ	ɯ
		[back]	−	+	+
		[labial]	−	−	+
		[lip protrusion]	−	−	−
	b.	Korean	i	ɨ	u
		[back]	−	+	+
		[labial]	−	−	+
		[lip protrusion]	−	−	+

6. See Note 2 for a brief discussion of the notions of marking statement, prohibition and repairs in the model of phonology adopted here.

Therefore, in borrowing Japanese words, Koreans repairs the illicit configuration [+labial, –lip protrusion] of Japanese by delinking [–lip protrusion] and inserting the unmarked [+lip protrusion], as in (9), where Lip is the articulator node from which the two features [labial] and [lip protrusion] are dependent:

(9)



In the case of the post-sibilant [ɨ] allophone of Japanese, we must assume that it is [–labial] as shown in (8).[7] Therefore this allophone is featurally identical to the Korean [ɨ].

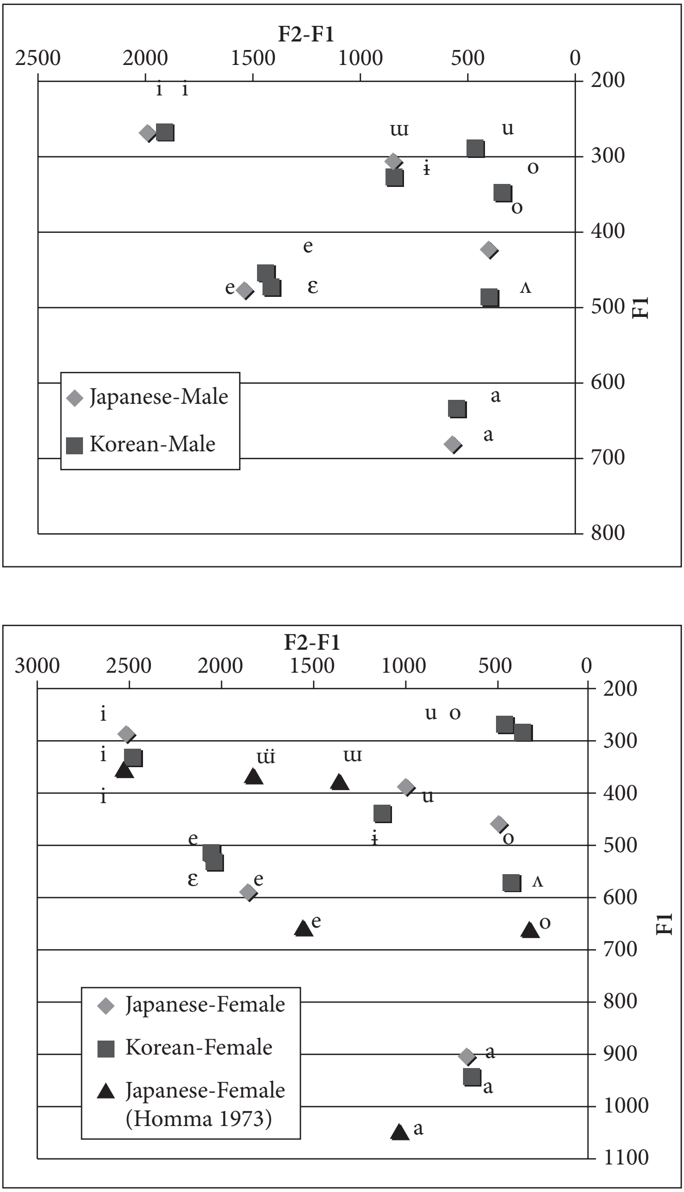
Ito, Kang and Kenstowicz show that the Korean adaptations of Japanese vowels are problematic for the two models of loanword adaptation discussed above. The fact that the Korean adaptation takes account of the Japanese [ɨ] allophone is problematic for the phonological model of LaCharité and Paradis (2005). This segment would not be expected to be present at the phonemic level, where the loanword phonology operates in their model. Yet it is precisely in the post-sibilant context that Japanese /ɯ/ is adapted as Korean /ɨ/, strongly suggesting that the adaptation is taking account of this predictable allophone.[8]

A model of loanword adaptation that assumes that mapping is strictly based on phonetic similarity between the outputs of the donor and recipient languages (Silverman 1992; cf. Peperkamp 2002; Peperkamp & Dupoux 2003) would also fail to provide a straightforward account for the adaptation of Japanese /ɯ/. When we examine the acoustic properties of these vowels, given the close proximity of the Japanese /i/, /e/, /a/, and /o/ and their Korean counterparts in acoustic space, the adaptation of these vowels can be accounted for in terms of acoustic similarity. However, this model incorrectly predicts that Japanese [ɯ] should be adapted as Korean [ɨ], and not [u], since they are the most similar in the acoustic map, as illustrated in (10) (figures from Ito, Kang & Kenstowicz (2007)).

7. Evidence for this is that in the Yonaguni dialect of Okinawa (Joo 1977, p. 125) /u/ after /s/, /z/, /t/ and /d/ has merged with /i/ (see Ito, Kang & Kenstowicz (2006)).

8. Note that /u/ occurs without problems after the sibilants in Korean (e.g. *supak* 'watermelon'), thus precluding an independently motivated adjustment /u/ → [i] after sibilants.

(10)



No matter how the various formants of the high vowels are weighted, as Ito, Kang and Kenstowicz observe, Japanese [ɯ] is best matched by Korean /ɨ/ in purely acoustic terms. Nevertheless, in loanword adaptation Japanese [ɯ] is adapted as Korean /u/ except after sibilants.

In light of these observations, Ito, Kang and Kenstowicz argue for a third model of loanword adaptation: This model assumes that enough phonetic detail must be retrieved from the donor language so as to distinguish the two allophonic variants of Japanese /ɯ/. It follows that the input to the adaptation must be a surface phonetic representation, At the same time, this model also assumes that the adaptations operate on abstract featural representations of the source language: they involve involve phonological operations on features.

4. Evidence for loanword adaptations in perception

Peperkamp and Dupoux (2003) review psycholinguistic evidence showing that all aspects of non-native phonological structure, including segments, prosodic, and syllable phonotactics, are systematically distorted during speech perception; i.e. non-native sound structures are adjusted both by monolinguals and by bilinguals. Comparing loanword adaptations to experimental speech perception data, they point to a number of striking correspondences. For instance, Korean listeners find it hard to distinguish between the English consonants [r] and [l] in CV stimuli (Ingram & See-Gyoon 1998), and in English loanwords word-initial [l] is adapted as [r] (Kenstowicz & Sohn 2001). In a similar vein, French listeners have severe difficulties perceiving stress contrasts (Dupoux et al. 1997) and in loanwords, stress is systematically word-final, regardless of the position of stress in the source word.

A striking case which demonstrates the fundamental role played by speech perception in the nativization of loanwords is the perception of illusory vowels in consonant clusters by Japanese speakers. These individuals perceive illusory epenthetic vowels in sequences of segments that do not fit the syllable structure of their native language. Moreover, Japanese speakers often insert epenthetic vowels when they pronounce loanwords involving these same clusters:

(11)	a.	[ma.kɯ.do.na.rɯ.do]	'MacDonald' (Japanese)
	b.	[sɯ.to.ra.i.ko]	'strike' (Japanese)

At this point, an obvious question arises: Are such epenthetic vowels inserted in production or perception? In a series of behavioral experiments, Dupoux and colleagues (Dupoux et al. 1999; Dupoux et al. 2001; Dehaene-Lambertz et al. 2000) compared Japanese listeners with French listeners in their perception of consonant clusters. For instance, Dupoux et al. (1999), give an off-line phoneme detection task (Experiment 1) in which they used a series of six items created from naturally produced nonce words (e.g. [abuno], [akumo], [ebuzo], [egudo], etc.) in which they gradually reduced the duration of the vowel [u] to zero milliseconds. While listening to a recording of the sounds, participants were asked if the item they heard contained the sound [u]. Japanese listeners, unlike French listeners, overwhelmingly judged that the vowel was present at all levels of vowel length. Strikingly, this was the case seventy percent of the time even when the vowel had been completely removed. The French participants, on the other hand, judged that the vowel was absent in the no-vowel condition about 90% of the time and that a vowel was present only in 50% of the intermediate cases. These results were confirmed in other experiments, which have led Dupoux and colleagues to conclude that the influence of native language phonotactics can be so robust that listeners perceive illusory vowels to accommodate illicit sequences of segments in their L1.

Similarly, Kabak and Idsardi (2006) show that Korean listeners perceive illusory vowels within consonantal clusters that are illicit in that syllable structure. As was observed for Japanese, Korean speakers also insert epenthetic vowels within the same clusters. (cf. [a.i.sɯ.kʰɯ.rim] 'ice cream', [khɯ.ri.sɯ.ma.sɯ] 'Christmas').

Given the overall similarity between speech perception data and loanword adaptations, Peperkamp and Dupoux (2003) propose that all loanword adaptations apply in perception. Though this is too strong, loanwords with adaptations can appear in situations of bilingualism as in the North End/Italian example discussed previously. However, it seems that many or even most borrowings do not occur in a situation of bilingualism, but in a situation of language contact between monolingual or imperfectly bilingual speakers. Here the role of perception and learning is fundamental. Observe that, as discussed above, for loanwords that develop in this situation, the input is a surface phonetic representation. In this paper I will focus on such borrowings.

Peperkamp and Dupoux (*ibid*) also assume that the adaptations observed in perception involve phonetically minimal transformations. That being said, it is unclear how adaptations such as epenthesis or stress shifts, are phonetically minimal. Furthermore, the adaptations discussed by Ito, Kang and Kenstowicz which have surface phonetic representation as inputs, and therefore must have occurred in perception, are not phonetically minimal, but actually appear to involve clear phonological operations.

In this paper, I will follow Peperkamp and Dupoux (*ibid*) in assuming that at least some, if not most, of the phonological adaptations characterizing loanword phonology occur during perception. Moreover, I will also propose that these adaptations involve the same phonological processes that characterize speech production.[9]

Before going on, however, I would like to address a fundamental problem of a philosophical nature: if perception involves production, or better construction of a representation through production, we would be experiencing only fallible

9. See Kim (this volume) and Boerma and Hamann (this volume) for a similar approach to loanword phonology.

representations – illusions. We would be detached from the hardness of reality. This cannot be correct. We do experience reality directly, or in the case of speech, we do have a direct access to the acoustic reality of the message. I will try to address this issue in the next section.

5. Perception

As first pointed out by Immanuel Kant in the 18th century, perception is not neutral, or passive, but rather involves a complex inferential computation by which sensory data from a given object are organized, categorized and adjusted by accessing abstract cognitive categories and previous knowledge.

At the same time, perception must be anchored to reality. As observed by Gibson (1979) and Merleau-Ponty (1945), perception must be tied to our being in the environment. Individuals must also have a direct, unmediated experience of their surroundings so that they can interact with it properly. Perception through inference is obviously indirect and prone to mistakes and does not capture the concrete experience of the world obtained through perceiving bodies immersed in the moving texture of reality.

Following Norman (2001), I will propose that these two aspects of perception correspond to two different but interacting perceptual systems: the "ventral" system dedicated to the identification and recognition of objects and events in the environment and the "dorsal" system dedicated to picking up information for a proper interaction with the reality. In the first system perception is indirect, mediated by the cognitive system and memory; in the second system it is direct, immediately given in the process of sensory-motor integration.

Recent work in the cortical organization of vision has emphasized that sensory input must interface both with conceptual systems (for object recognition) and with motor systems (e.g. visually guided reaching/grasping) (Ungerleider & Mishkin1982; Milner & Goodale 1995).

It has been demonstrated empirically that these two interface systems comprise functionally and anatomically differentiated processing streams in which a ventral (occipital-temporal) stream supports object recognition/understanding (the "what" pathway), and a dorsal (occipital-parietal) stream supports visual-motor integration functions (the "where" pathway).

The primary function of the ventral system is the recognition and identification of objects and events in one's environment (Norman 2001). It compares visual inputs to stored information in an attempt to achieve a meaningful interpretation of those inputs. The ventral system deals mainly with the utilization of visual information for interpreting one's environment, The recognition and identification processes that are part of this interpretation require comparisons with stored representations. This system is therefore memory-based, using stored representations to recognize and identify objects and events.

In contrast, the primary function of the dorsal system is the analysis of visual inputs from the ambient array in order to allow interaction with the environment and the objects in it e.g. pointing, reaching, grasping, walking towards or through something, climbing, etc. (Norman 2001). The dorsal system picks up visual information to allow the organism to function in the environment. It is a system that picks up invariants in the ambient array by directly "resonating"—to use Gibson's (1979) ecological terminology — to the features to that array. It is a system that has a direct relationship with the environment. It anchors the perceiving individual to the external reality. Some have suggested that all the information pick up for the performance of well-ingrained actions or behaviors is implemented by the dorsal system. The dorsal visual stream is thus particularly geared for visual-motor integration, as required in visually guided reaching and orienting responses (Rizzolati, Fogassi & Gallese 1997).

The dorsal system deals mainly with the utilization of visual information for the guidance of behavior in one's environment. Much of our day-to-day pick up of visual information is carried out by the dorsal system without much conscious awareness. As Norman (2001) points out, this system is ecological in the sense of Gibson (1979). It directly picks up information from the ambient array for or through action. From now on, I will refer to the outputs of this perceptual system as "*sensory intuitions*" insofar as they have a unmediated, direct, or "naive" relation to objective reality as in Kantian empirical intuitions.

The ventral system is instead, a "higher" system that deals with the interface between the visual input and cognition, and we are normally conscious of its output. It is the system that tries to make sense of situations. This is achieved via an indirect, post-sensory, inferential nature. Interpretation is an inextricable part of the perception processes characteristic of the ventral system. This is the system that implements the type of indirect, inferential perception advocated by constructivist models of perception (see Helmholtz 1867; Rock 1977, 1983, 1997; Gregory 1993; Boring 1946; Epstein 1982; see also Piaget (1937, 1969); Marr (1982); Anderson (1985); Kanisza (1979).[10] I will refer to the outputs of this perceptual process with

10. In this way, following Van Leuween (1998) and Norman (2001), the Gibsonian approach on perception, in which perception is simply the pick up of information from invariants in the ambient environment, and which by itself, is inadequate as a general model of perception, can be integrated with a more adequate constructivist model of perception.

the term "*apprehensions*". In the terminology adopted here, visual perception thus includes both "sensory intuitions" and "apprehensions."

6. Speech perception

Following Hickock and Poeppel (2004), I assume that two similar perceptual systems are also active in speech perception. But in order not to create confusion between visual perception and speech perception, this text will refer to the linguistic ventral system **as** the **top-down system**, and the dorsal system as the **bottom-up system**.

The objective of listeners during the act of speech perception is to access the meaning of the utterance they hear, this can be accomplished only via the identification of the vocabulary items (morphemes/words) used in it and the interpretation of their structural significance in the syntactic environment in which they occur. Crucially this identification/recognition depends on the identification of the sounds comprising the vocabulary items. In classical semiological terms, one must identify the *significans* (the form of the expression) in the utterance to access its *significatum*.

Given that identification and recognition are characteristic functions of the top-down (ventral) system, one could propose that speech perception is implemented only by the top-down system. However I propose that the bottom-up (dorsal) system also has a role insofar as it picks up information and provides it to the top-down system for analysis. Thus, the bottom-up system picks up acoustic data and converts them into acoustic representations —"sensory intuitions"— that are provided to the top-down system which converts them into articulatory representations — a process of "apprehension". These articulatory representations can be used to identify and recognize the vocabulary items, which, as discussed in Section 1, are represented in terms of articulatory features. Thus, whereas the dorsal system in vision functions as a parallel perceptual system dedicated to visual-motor integration, in the case of speech perception where the main goal is identification and recognition of *significanta* in the acoustic input, a goal that can be achieved only by the top-down system, the bottom- up system must pick up acoustic information from the input and produce "sensory intuitions" or the top-down system where these representations are interpreted/apprehended.[11]

---

11. But note that later, following Hickock and Poeppel (2004) I will propose that there is a direct path between acoustic representations in echoic memory and meaning/concepts, therefore a speech perception stream that by-passes the top-down system. This path between

Following Blumstein and Stevens (1981), Stevens and Blumstein (1981), and Stevens (1998), I assume that the acoustic signal is analyzed to extract invariant acoustic properties or patterns which excite designated acoustic detectors in the bottom-up system. An acoustic representation is produced through this parsing of the acoustic signal. As proposed above, this retrieval of linguistic information from the acoustic signal is implemented by the bottom-up system.

The point is that in contrast to the discretely specified phonological representation of an utterance, the acoustic signal that is produced by a speaker is continuous. It is an analog signal that is generated by continuous movements of a set of articulatory and respiratory structures. However, as mentioned in Section 1, the relations between the articulatory and the acoustic representations of speech have certain quasi-discrete or quantal characteristics that are exploited in speech production (Stevens 1972, 1989). These quantal attributes help to simplify the process of uncovering the discretely specified segmental and categorical representations that are building blocks of words (Stevens 2002).

As proposed by Stevens (2002), the retrieval of linguistic information from the acoustic signal proceeds as follows. First, the locations and types of the basic acoustic landmarks in the signal are established. These acoustic landmarks are identified by the locations of low-frequency energy peaks, energy minima with no acoustic discontinuities, and particular types of abrupt acoustic events. From these acoustic landmarks certain articulatory events can be hypothesized: the speaker produced a maximum opening in the oral cavity, or a minimum opening without an abrupt acoustic discontinuity, or a narrowing in the oral cavity sufficient to create several types of acoustic discontinuity (Stevens 2002). Such landmarks provide evidence for distinctive features such as [consonant], [sonorant], [continuant], [strident], the so-called articulator-free features in Halle (1995). The second step consists of the extraction of acoustic cues from the signal in the vicinity of the landmarks. These cues are derived by first measuring the time course of certain acoustic parameters such as the frequencies of spectral peaks or spectrum amplitudes, in particular frequency ranges, and then specifying particular attributes of these parameter tracks (Stevens 2002). These acoustic parameters provide evidence that articulators are involved in producing the landmarks and demonstrate how these articulators are positioned and shaped.

However, coarticulation and other contextual and prosodic adjustments affect cues and landmarks reducing their strength or just eliminating them. Thus, uncovering of the segments and features that underlie the words in an utterance, then,

"acoustic" *significantia* and *significata*, however, is possible only under special circumstances, namely an established familiarity of the *significans*.

involves using acoustic data to make inferences about the gestures that the speaker uses to implement these features, since the gestures tend to bear a closer relation to the features than do the acoustic patterns, as first pointed out by Liberman (1956) and discussed in Section (1). This inferential activity is part of the top-down system. Stevens (2002) proposes that once the sequence of landmarks has been identified and a set of acoustic cues has been evaluated in the vicinity of each landmark, the next step is to convert this landmark/cue pattern into a symbolic or quantal description consisting of a sequence of feature bundles. This conversion is carried out in a specialized phonological module, which, according to my proposal, forms part of the top-down system. This module consists of a set of submodules, one for each feature. Every submodule examines the acoustic landmarks and cues that are relevant to the feature and analyzes them in term of their overall environment and prosodic context. It then produces an inference concerning the specification of the feature. The details of each of these submodules, including the selection and design of acoustic cues that form possible inputs for each module, are beyond the scope of this article (see Stevens (2002) for more discussion).

Through the activity of the phonological module, the acoustic signal of an utterance is converted into an array of articulatory features. I assume here that each submodule interprets the acoustic data autonomously from the other submodules. It is the duty of a special component called the synthesis component, to read out all of the articulatory information provided by these submodules to construct an underlying representation which is necessary to access the meaning of the utterance and then derive from it the surface representation which is behind the utterance that produced the acoustic input. Recognition/identification/apprehension of the acoustic input consists of this process of interpretation and construction (see Section 10 for some hypotheses on how this occurs). One of the first steps in the synthesis component, after the piecing together of the information provided by the phonetic module, is the identification of the morphemes and words present in them by matching them with morphemes and words that are stored in a long term memory dictionary as arrays of features. This matching process is then followed by a top-down construction of a sentence which attempts to make sense of the morphemes and words that have been activated in long-term memory. I propose that speech perception thus also involves activation of the production component of the grammar that implements the computation involved in this construction (see Section 10 for more detail).

I assume that the mapping into articulatory representations, the analysis and top-down construction of the sentence discussed above are implemented in the verbal working memory, the so called phonological loop (Baddelley 1992). It is here that the analysis-by-synthesis of the sentence is implemented. Hickok and Poeppel (2004) argue that verbal working memory relies on a auditory–motor integration network (Aboitiz & García 1997). Indeed, verbal working memory (and perhaps working memory in general) can be viewed as a form of sensory– motor integration (Wilson 2001). For example, in Baddeley's model (1992), the phonological loop is essentially a mechanism for using motor systems (articulatory rehearsal) to keep sensory-based representations (the phonological store) active. Hickok and Poeppel postulate an explicit neural network for verbal working memory. In their model, the phonological store is identified with the Superior Temporal Gyrus systems supporting acoustically-based representations of speech which the articulatory rehearsal component maps onto frontal systems supporting articulatorily-based representations of speech. Evidence from Awh et al. (1996) indicates that the articulatory rehearsal component involves left frontal cortices, notably portions of Broca's area and more dorsal pre-motor regions.

Notice that a working memory phonological buffer (the phonological loop of Baddeley (1992)) is fundamental to understanding language acquisition (Doupe & Kuhl 1999): in fact, for the child to learn to articulate the speech sounds in his or her linguistic environment, there must exist the following components: (i) a mechanism by which sensory representations of speech uttered by others can be stored (echoic memory); (ii) a mechanism by which the sensory input is analyzed and converted into an articulatory representation (the analysis-by-synthesis component of the phonological buffer (see Section 10), (iii) a mechanism by which the child's articulatory attempts can be compared against these stored representations, and by which the degree of mismatch revealed by this comparison can be used to shape future articulatory attempts (the "comparator" in the top-down system (see Section 10)). Although such a network obviously assumes less importance in adult speakers, the fact that new words from one's own language and from foreign languages can be learned at any age show that it continues to operate throughout life.[12]

We expect the motor system to be active in this working memory component. And in fact, recent neurological studies have shown that perceiving speech involves neural activity of the mirror neurons and the motor system. The mirror neurons are a particular class of neurons that exhibit excitations not only when an individual executes a particular action but also when the same individual observes the action being executed by another individual. The existence of these neurons (see Di Pellegrino, Fadiga, Fogassi, Gallese & Rizzolati 1992) provides direct neural evidence for motor system involvement in perception. (Rizzolati & Craighero 2004;

12. Further evidence that the phonological loop plays a role in adults is provided from articulatory decline following late-onset deafness (Waldstein, 1989), from the effects of delayed auditory feedback on speech articulation (Yates, 1963), and from altered speech feedback experiments (Houde & Jordan, 1998).

Rizzolati, Fogassi & Gallese 2001). In Rizzolati and Arbib's (1998) words, "taken together, the human and monkey data indicate that, in primates, there is a fundamental mechanism for action recognition. … Individuals recognize actions made by others because the neural pattern elicited in their premotor areas during action observation is similar to that internally generated to produce that action." (p. 190)[13]

Two recent studies involving the use of transcranial magnetic stimulation of the motor cortex have demonstrated activation of speech-related muscles during speech perception. Fadiga et al. (2002) found that when listeners hear utterances that include lingual consonants, they show enhanced muscle activity in the tongue. Watkins and colleagues (Watkins, Strafella & Paus 2003) found that both while listening to speech and while seeing speech-related lip movements, people show enhanced muscle activity in the lips. Complimentarily, two fMRI studies (Pulvermüller et al. 2006; Wilson, Saygin, Sereno & Iacoboni 2004) demonstrated that there is overlap between the cortical areas during speech production and those active during passive listening to speech. As shown by Fadiga et al. (2002), the same motor centers in the brain are active both in the production of speech and in speech perception, where the perceiver engages in no overt motor activity.[14, 15]

---

13. Recently researchers have shown that Rizzolati and Arbib's (1998) "fundamental mechanism for action recognition" also has ties with general audition. Kohler et al. (2002) found neurons in the pre-motor cortex of monkeys that respond not only when the monkey performs a specific action (e.g. breaking a nut) or sees the action performed by someone else, but also when the monkey merely hears the sound that is caused by the specific action (e.g. the cracking noise of the nutshell being broken).

14. Still, as discussed below, it is possible to have basic speech perception, in the sense of access to meaning of an utterance, without accessing the top-down system and the motor systems.

15. Fadiga, along the lines of the motor theory of speech perception proposed by Liberman and Mattingly (1985, 1989), or better in terms of the Direct Realism model of Fowler (1986, 1994, 1996), states that "speech perception and speech production processes use a common repertoire of motor primitives that during speech production are at the basis of articulatory gesture generation, while during speech perception are activated in the listener as the result of an acoustically evoked motor 'resonance'" (Fadiga et al. 2002). Fadiga, therefore, assumes that there is direct perceptual relation between acoustic stimuli and motor activity, which he calls an acoustically evoked motor 'resonance', as in the ecological theory of perception of Gibson (1979). But how this resonance is implemented is unclear. Infact, at the neural activity of the motor and premotor systems does not need to be seen in terms of Gibsonian 'resonance' but could be seen rather in a more indirect, memory-mediated manner in which these neural activations are brought about by the phonological loop, as discussed above.

7. Phonological adaptations in perception

In the normal situations of speech interaction, the goal of speech perception is comprehension, the identification of the *significatum*, the meaning carried by the *significans* of the utterance. Identification of the *significans* is fundamental in the learning of new words regardless of whether the new word belongs to the learner's native language. In either case, a mental representation of the *significans* of the word to be committed to memory must be constructed. This representation must include both phonological and morphological information. In order to be committed to memory, however, the *significans* must be interpretable, i.e. identified and recognized by the cognitive systems Therefore, if it contains unfamiliar or illicit, configurations, i.e. uninterpretable configurations, they must be adjusted so to be identified.

When faced with an unfamiliar linguistic sound, a perceiver has an obvious problem insofar as a configuration that is uninterpretable in terms of his own system of linguistic knowledge must be analyzed in terms of this system. A first rough account of what happens in this case is the following. If a segment, or a syllabic combination of segments, is unfamiliar, foreign, i.e. absent fom L1, a speaker has no instructions for how to produce it, i.e. no representation of it with the right combinations of features, or segments in the case of syllable configurations. In particular, as proposed in Calabrese (2005) (see also Footnote 2), the absence of a segment, or a syllabic combination of segments, indicate the absence, or unavailability, of computational programs coordinating their featural configurations into articulatory commands. The absence of such a program is formalized as an active constraint against the configurations. Configurations violating an active constraint are repaired. The repair that occur in this case then indicate the featural/configurational manipulations that adjust the representations and make their conversion into articulatory commands possible (see Calabrese (2005) for more discussion of this model). As proposed in Calabrese (2005), active constraints are checked throughout the derivation, and if violated, repairs apply.

If production plays a role in perception, then it follows that active constraints and repairs play a role as well. Accordingly, unfamiliar sounds disallowed by active constraints must be repaired in perceptual representations, thereby resulting in a perceptual adaptation of the unfamiliar sounds.

At this point it is useful to focus a hypothetical instance of this process. The event begins with the production of an utterance by a foreigner in his or her own language. The native listener does not speak this language well. Therefore, this utterance contains sound configurations that are novel to him/her and are disallowed by an active constraint in his grammar. This utterance produces a certain acoustic signal, which enters the bottom-up perceptual system and is recorded in the acoustic short-term memory storage, the short-term echoic memory. In the acoustic analysis component, this signal is analyzed in terms of its invariant acoustic properties, as discussed in the preceding section, but it is still an uninterpreted "sensory intuition". Thereafter, this acoustic information is sent to the top-down perceptual system for identification/recognition. The phonological module converts it into articulatory information by converting the acoustic invariant properties present in the acoustic data into articulatory features/structures. Remember that each submodule of the phonological module interprets the acoustic data autonomously from the other submodules and that the synthesis component of the phonological working memory buffer uses the pieces of articulatory information (feature specifications, prosodic and syllabic articulatory cues, etc.) provided by the submodules to construct an underlying representation and then derive from it a surface representation. If the acoustic input contained new or unfamiliar sounds or sound configurations, the synthesis component cannot put together the pieces and cues provided by the phonological module in the case of these sounds, or sound configurations. In fact, if these pieces and cues were put together as indicated in the acoustic input, they would form illicit, "unpronounceable" configurations blocked by active constraints of the grammar. These illicit configurations must then be adjusted and repaired. The application of these repair operations will produce a "more familiar" nativized representation.[16]

Consider a concrete example. Take an Italian listener like me. The Italian vowel system does not have the [+low, −back] vowel /æ/ of the English word /kæt/. In the terminology developed above, this means that there is an active constraint disallowing the feature configuration [+low, −back]. The acoustic input of this vowel will be characterized by a low first formant and a higher second formant. The submodule for the feature [low] detects the first acoustic property and assigns the specification [+low] to the articulatory representation of the word /kæt/ that is being built in the synthesis component. The submodule for the feature [back] assigns to it the feature specification [−back]. The information provided by the acoustic input through the phonological module also requires the simultaneous articulation of these two features in the vowel. However, the synthesis component operates according to the grammar of L1. Thus the fact that there is an active constraint disallowing this feature combination in L1 prevents the synthesis component to put together these two feature specifications into the same feature bundle. Thus, this feature combination cannot be created in the mapping from the acoustic invariant properties of the signal into the articulatory representation of the synthesis

16. See Kim (this volume) for a model in which acoustic cues for L2 sounds are mapped into L1 featural representations without repair operations.

component insofar as it is illicit in Italian. This featural configuration must then be repaired; it may either by changing the feature [+low] into [−low], deriving [ɛ] from /æ/, or by changing the feature [−back] into [+back], deriving [a] from /æ/. It follows that I may interpret/apprehend /kæt/ as [kɛt] or as [kat]. Accordingly, the acoustic configuration characterizing this vowel (with its low first formant and higher second formant) cannot be associated with the appropriate articulatory feature configuration ([+low, −back]), resulting in a "phonetic illusion".[17]

17. The input to the analysis and interpretation in linguistic perception is the word in its surface representation. This is how the L2 word is "heard". Assuming a general principle of economy of operations is executed during perception, just as in production, as proposed in Calabrese (2005), only the configurations of the L2 word that are illicit in L1 must be repaired, i.e. and not those that are licit in L1. Only the minimum that is necessary to fix the input is changed, the rest must be preserved. The preservation of the licit aspects of the input may explain why the treatment of L2 words in L1 often involve processes that are not part of L1 phonology. One such discrepancy between processes in native and loanword phonology is provided by the treatment of English loanwords into Korean. In Korean, [s] is not allowed in syllable codas. In the native phonology, an underlying /s/ is realized as [t] when it occurs in coda position (ia). However, in English loanwords, words with [s] in coda position systematically undergo epenthesis (ib) (Kenstowicz & Sohn 2001).

(i)	a.	/nas/ [nat] 'sickle-NOM'		
		/nas + ɨl/		[nasɨl]
	b.	[posɨ]	<	'boss'
		[kɨrasɨ]	<	'glass'
		[mausɨ]	<	'mouse'
		[k^harisɨma]	<	'charisma'

The reason for the selection of epenthesis instead of neutralization to remove the coda consonant results from the desire to preserve the featural composition of [s] that is otherwise licit in the language (see Boersma & Hamann (this volume) for an account of these facts in a framework similar to the one proposed here but adopting OT).

However, not all cases of a discrepancy between processes in native and loanword phonology can be accounted for in a similar way. Consider, for illustration, the treatment of loanwords from French into Fula. In the latter language, neither onset nor coda clusters are permitted. In loanwords from French, an epenthetic vowel is added after the second consonant in liquid+obstruent clusters see (iiai), but between the consonants of obstruent+liquid clusters see (iiaii) (Paradis & LaCharité 1997). In the native phonology, however, the epenthetic vowel is always inserted after the second consonant, both in the case of liquid+obstruent clusters, as in (iibi), and in the much rarer case of obstruent+liquid clusters, see (iibii) (data from Paradis (1992)).

(ii)	a.	i.	[karda]	<	Fr. carde [kard] 'card (comb)'
			[fɔrsɔ]	<	Fr. force [fɔrs] 'force'
		ii.	[kalaːs]	<	Fr. classe[klas] 'flag'

At this juncture it is important to deal with a common aspect of the speaker/listener's experience of foreign sound that is aptly described in the following anecdote (from Jacobs & Gussenhoven (2000: 203)): "Valdman (1973) reports the reaction by a Haitian-Creole-speaking maid who attended evening literacy classes to her teacher's pronunciation of oeuf 'egg' as [ze]: she decided to leave the class. Although she herself pronounced it that way, she was aware that her bilingual employers realized it as [zø]". In other words, she did not know how to pronounce the word, but she did know how it sounded. A listener can be aware of the acoustic shape of a given sound, although his own pronunciation is different.[18]

The simplest hypothesis to account for a situation like the one above is that upon presentation of a foreign word/sounds, a learner/listener learns perfectly faithful representations of them without access to the production system which we know is impaired in lacking relevant articulatory instructions for the foreign sounds. These faithful underlying representations would be modified by the speakers when they are pronounced during production; at the same time they could be used as criteria to judge the difference between their correct forms in L2 and their adapted forms in the speaker pronunciation, as in the above event. This would require a perceptual system that is independent of the grammatical system which faithfully converts acoustic inputs into underlying featural representations (Jacobs & Gussenhoven (200); Hale & Reiss (2000)).

The evidence provided earlier that the perceptual representations of listeners are distorted by their native grammar (Dupoux et al. 1999; Dupoux, et al. 2001; Dehaene-Lambertz et al. 2000) shows that this hypothesis cannot be maintained. Learning the exponent of a new or foreign word means learning the articulatory patterns expressed in its featural composition. If there are grammatical constraints against the featural organization of these patterns, the word cannot be learned as such and its featural organization must be adjusted, as discussed above. Therefore

b. i. /talk+ru/ → [talkuru] 'amulet'
ii. /sokl+ka/ → [soklaka] 'need'

It is unclear to me how preservation of the licit featural configurations of the input can account for the difference between native and loanword phonology in this case. Such cases are perhaps better analyzed by hypothesizing a special loanword phonology grammatical component. This component would include the active marking statements of the native language and the related repair processes that were grammaticalized/istitutionalized during the contact with the foreign language. Further research is required to establish if such a special grammatical component is needed.

18. The same occurs in children. When my daughter was about two year old, she used to pronounce [ʃ] as [s]; therefore, she said [sIp] instead of [ʃIp] 'ship'. So once I tested her and, while pointing to the picture of a ship, I asked: "Is this a [sIp]?". She replied, "No! it is a [sIp]!"

it follows that a faithful conversion of a foreign word into a long-term memory underlying representation is impossible.

Furthermore, Section 9 will address evidence that children older than 5 or 6 and adults are "behaviorally deafened" to foreign contrasts not shared with the native language. This further weakens the idea that it is possible to convert a foreign linguistic input into a faithful featural representation. It is true, as will soon be evident, that by focusing attention on the acoustic input and intensive training, this "deafness" can be overcome. But this training involves articulatory exercises that teach the learner to acquire the articulatory pattern that are constrained in his L1 grammar. Once these articulatory patterns are acquired, the L2 sound is learned and its accurate "perception" becomes possible. But this is precisely what is argued here. An accurate perception of a sound requires learning how to produce that sound articulatorily.

However, if what is perceived is simply identical to what is produced – as can be concluded at first from the above proposal, awareness of the acoustic shape of the given sound by a listener, when his own pronunciation of this sound is different, should be impossible. A more careful consideration of the theory shows that this consequence is incorrect. The bottom-up system in fact must include an echoic short-term memory (Neisser 1967) where aural acoustic representations[19] of speech sounds are preserved. These are the acoustic representations that are analyzed in the preliminary phase of speech perception.[20] We can also assume that such representations can be stored in long- term echoic memory with all other non-linguistic sounds.[21] Observe that, in the preliminary phase of speech analysis, where acoustic patterns are analyzed in terms of their invariant properties, there must be a basic ability to distinguish sounds so as to extract their characterizing features. It follows that we must be able to distinguish aural representations

19. The term *image* is often used in this case "aural acoustic image". I prefer to use representation to indicate that there is always a degree of symbolic conversion between the sound in itself and the sensory *representation* of it provided by our neural networks.

20. The output of this analysis must already be somewhat abstract insofar as linguistically irrelevant properties, such as the voice characteristics of the speaker that uttered the word, the rate of speech, distortions caused by a cold or sore throat and so on, can be neglected in the formation of the memorized linguistic representation of the sound. Obviously an issue here is how we store words in memory, be they as word-type or word-tokens (Goldstone & Kersten 2003; Hintzman 1986; Goldinger 1986, 1988). This issue falls outside the aegis of this article.

21. Anyone that has ever played with a short-wave radio knows that it is possible to understand if the language of the tuned station is Chinese, Russian, Arabic, etc. without knowing these languages – this is possible by accessing an aural memory of how those languages are spoken.

of sounds preserved in long-term echoic memory, and use them for comparison with items stored in the short-term echoic memory. It is therefore in the echoic long term memory of the Haitian maid that the standard sound [ø] for [zø] 'egg' was stored. Thus, although she could not pronounce it, she could use that echoic memory representation for comparisons.

Neurolinguistic evidence suggests that echoic memory is located in the auditory cortex (Näätänen 2001). If this is correct, one expects evidence of analytic activity in this part of the cortex though the fact is that several studies suggest that more complex events such as stream segregation – extracting the abstract sound patterns and invariant sound relationships – and categorical speech perception guided by language-specific memory traces may take place preattentively in the auditory cortex (Sussman et al. 1998, 1999; Tervaniemi et al. 1994; Paavilainen et al. 1999, 2001; Dehaene-Lambertz 1997; Phillips et al. 2000; Phillips 2001; Shestakova et al. 2002).

Distinguishing aural images of sounds, however, does not mean being able to recognize or identify them. Recognition and identification of a string of sounds involves determining how the feature specifications detected in the preliminary parse of the utterance acoustic input are organized into feature bundles and how these bundles are distributed in a syllabic string, i.e. it means identifying how these feature specifications are combined in the featural configurations of the string. To put this in Kantian terminology, we can say that only at this point could we have perceptual judgments and thus state whether or not a certain sound is /ø/ or /æ/, i.e. a feature bundle containing the configuration [−consonantal, −back, +round], or a feature bundle containing the configuration [−consonantal, +low, −back]. Insofar as the lack of relevant articulatory instructions block these feature configurations, the possibility of these judgments is prevented, and in so doing thus the recognition and identification of the relevant sounds is not possible.

It follows that a listener can be aware of distinctions among unfamiliar sounds without being able to identify them – this awareness is possible because of acoustic images of the sounds stored in echoic memory. However, although this listener may feel that these sounds are different, he cannot know how they are different, and when forced to represent them in the construction of the linguistic representation in apprehension, he must adjust them into more familiar configurations.

The problems pointed out by Ito, Kang and Kenstowicz (2006) are then easily addressed. The input to perception is a surface phonetic representation. If an allophonic property is licit and can therefore be interpreted it will appear in the loan. Thus Japanese back unround vowel [ɨ] found after sibilants, an allophone, can be adopted as such in Korean loanwords insofar as such a vowel is present in the Korean phonological system.

Furthermore, given that acoustic proximity does not play a role in the model proposed here, the Japanese back vowels with the configuration [+labial, –lip compression] will be repaired by Korean speakers as discussed in Section 3.3.

Japanese vowel epenthesis may be accounted for as follows: Japanese disallows complex onsets and complex codas. Simple codas are restricted to the first member of a geminate, or to a nasal glide. When presented with a word with such illicit clusters in the phonological working memory buffer, a Japanese speaker cannot construct a mental representation of it because of the active constraints active in his grammar. The listener resorts to epenthesis to fix this problem. The presence of the illusory vowels in perception then occurs.

Just as in the nativization of the high [–ATR] vowels of English to high [+ATR] vowels in loanwords to French, Spanish and Italian, I assume that we are dealing with a case of repair. The configuration [+high, –ATR] of English I, U is blocked by an active marking statement in these languages which is addressed by delinking the feature specification [–ATR] and replacing it with [+ATR]. This can occur both in perception or production as in all other cases discussed previously.

Following Dupoux et al. (1999), I used the term "phonetic illusion" to characterize the perceptual adjustments discussed above. But these are not "sensory illusions" in the sense that something is heard which is actually not part of the stimulus, rather, I assert that they are representational illusions which are generated by the linguistic operations that are used to construct adequate mental representations of linguistic objects. Given that these representations are memorized, the illusions are accessible only through recall. The Japanese listeners, in fact, report that there is a vowel in the consonantal clusters, but in that case they are "reporting" on the memorized representation of the stimulus, not on what they actually heard. To put this differently, they hear a certain acoustic input, say [eb—zo], as is but they cannot make sense of it perceptually because they lack this type of syllable structure in the production grammar. So they adjust the representation of this stimulus by epenthesizing a vowel [ebuzo]. It is on this memorized representation that they report, not on what they heard.

Once this is clear we have an account of the illusions found in the human experience of foreign speech sounds. Interestingly, these illusions are also observed in the experience of the native language, as most famously observed by Sapir (1933). While studying the Canadian Athabaskan language Sarcee, Sapir was puzzled by his informant John Whitney's insistence that there was a difference between *dini* 'this one' and *dini* 'it makes a sound' even though the two were phonetically homophonous to Sapir's trained ears. In order to explain his informant's intuition, Sapir postulated that the final vowels of words like *dini* make a sound with a "latent" consonant. Put differently, this suggests that there was another psychologically more accurate representation of the word that records the presence of this intuited sound: [dinit]. As Sapir observed, this was the *phonological representation* of the word.

Sapir called this experience by his informant a *phonetic illusion*. Two objectively identical stimuli *dini* were judged as different when associated with different phonological representations. Additional examples of this type of illusions are easy to find. Kenstowicz (1994), for example, reports that English speakers tend to perceive the intersyllabic consonantal material in *camper* and *anchor* as analogous to *clamber* and *anger*. This is an illusion, however. In most dialects (Malecot 1960) the nasal consonant is phonetically absent before such sounds as [p,t,k,s], so that *camper* and *anchor* have the same gross phonetic shape (C)ṼCVC (Ṽ a nasal vowel) as (C)VCVC *wrapper* and *acre*. While ṼCVC *anchor* belongs with VCVC *acre* phonetically, English speakers have the strong intuition associating it with VCCVC *anger*. This perceptual judgment is likewise based on the abstract phonological representations of these words.

In the model proposed here, these illusions are accounted for by assuming that the perception process involves access to the abstract phonological representation computed in the mind.[22]

8. Echoic memory and sensory intuitions

The echoic memory in the auditory cortex stores the acoustic features of the stimulus (Neisser 1967). The sensory information stored by echoic memory covers

---

22. Another model that assumes that adaptation of sounds in perception is due to the access to the articulation system is the Perceptual Assimilation Model of Best (1994, 1995). Best treats the adaptations found in the pronunciation of foreign sounds as assimilations to the closest native phoneme category on the basis of articulatory similarities and discrepancies.

Best assumes that listeners perceive information about articulatory gestures in the speech signal. Thus, they can perceive in non-native phones information about their gestural similarities to native phonemes. If the listener perceives the phones to be very similar in their articulatory-gestural properties to a native phoneme category, then the nonnative phones will be assimilated to this native phoneme category.

Besides the fact that this model rejects the use of phonological features for which there is overwhelming evidence, as discussed in Section 1, it remains unclear how information about the similarity of articulatory gestures in native and non-native phones is actually extracted from the acoustic signal without detecting that there are also acoustic discrepancies. It seems that there is a fundamental circularity in this approach to loanword adaptations. Notice furthermore that without recourse to phonological features the notion of gestural similarity become quite arbitrary. What makes an articulatory gesture more similar to a certain gesture than to another gesture?

Unfortunately, lack of space prevents a deeper discussion of this model.

all aspects of the stimulus.[23] I assume that echoic memory allows direct access to the acoustic input, it stores what I call the sensory intuition of the acoustic message, and involves the sensory system that closely tracks the acoustic stimuli. Experiments of duplex perception provide evidence for this system. Mann and Libermann (1983) conduct one such experiment in which the components of a syllable were presented dichotically. The base, presented to one ear, included steady-state formants for /a/ preceded by F1 and F2 transitions consistent with either /d/ or /g/. An F3 transition, presented to the other ear, distinguished /da/ from /ga/. Perception is called duplex because the transitions are heard in two different ways simultaneously. Listeners perceive a clear /da/ or a clear /ga/ in the ear receiving the base depending on which transition was the other ear was exposed to. In the ear receiving the transition, listeners hear a non-linguistic chirp. The listeners could be asked to attend the syllable or the chirps and make quite different judgments on them. They responded quite differently, even though the judgments were based on the same acoustic input. On the one end, this can be seen as evidence of a special system interpreting the combined sensory input of both ears and producing a linguistic percept; this is the Top-Down system advocated here. On the other end, it provides evidence of a system tracking the acoustic signal and storing it as such, a system referred to here as echoic memory, part of the bottom-up system.

There is also evidence that the auditory cortex maintains more permanent representations of the auditory past. Recent neurological evidence suggest that there is a set of sound memory traces, and of memory traces of sound combinations representing syllables and words, of the language involved in the left auditory cortex (Näätänen 2001; Pulvermüller et al. 2001; Pulvermüller et al. 2006) When a familiar speech sound is presented, it activates the corresponding phonetic trace or recognition model (in addition to the different sound-analysis mechanisms common to speech sounds and equally complex non-speech sounds).[24]

I propose that learning to speak a new language, in addition to learning how to articulate the different combination of features and syllabic organization which characterizes that particular language, also involves the formation of discrimination and recognition patterns for the acoustic counterparts of these features. This acoustic discrimination and recognition patterns help the further analysis and processing of the acoustic signal into featural representations by the top-down system.

---

23. Experimental evidence suggests that the memory-trace durations for echoic memory last between 5 and 10 seconds. Echoic memory can occur outside of conscious experience and attention-independently; (Woldorff, Hackley and Hillyard, 1991).

24. This auditory cortex process itself is *pre-perceptual*, but tends to trigger frontal cortex activity which probably underlies the initiation of attention switch to sound change.

Hickock and Poeppel (2004), furthermore, provide evidence for a cortical network which performs a direct mapping between acoustic representations and conceptual-semantic representations. Learning words, and any sublexical process requiring the analysis of phonological exponents requires the mapping of acoustic invariant properties of the signal including these words into articulatory feature representations and subsequent processing in the top-down system. But, once a word is learned, used, and heard many times, becoming thus totally familiar, it is obviously uneconomical to go through the phonological module and analysis by the top-down system every time that the same word is heard. One could assume, instead, that recognition of familiar and commonly used words and constructions may simply bypass analysis by the top-down system, and directly activate the exponents of the dictionary in the phonological working memory buffer through a direct association between acoustic representations stored in long-term memory echoic memory and word exponents in the dictionary. Hickock and Poeppel demonstrate that this must be the case by providing a variety of evidence, the most compelling of which is the dissociations we observe in aphasia patients. Namely, Broca aphasiacs demonstrate an ability for word recognition and comprehension despite widespread damage to the production system, and to what is known here as the phonological working memory buffer. Comparatively, Wernicke aphasiacs and Word deafness syndrome patients are characterized by their inability to access the dictionary, and a subsequent lack of word comprehension; on the other hand, both have an intact production system, and an ability to repeat syllables and words, which is explainable only if we also assume an intact phonological module and phonological working memory buffer.

9. Acoustic inputs and phonological discrimination

It is often stated that children become "deaf" to foreign phonological contrasts in the process of language acquisition.

Prior to the period of six to nine months of age, infants apparently can discriminate any sounds contrasts in any language. By the end of the first year, however, they apparently can no longer discriminate most sounds that do not contrast in the ambient language (Aslin, Jusczyk, & Pisoni 1998; Best, McRoberts, LaFleur, & Silver-Isenstadt 1995; Polka & Werker 1994; Werker & Lalonde 1989; Werker & Tees 1984). This developmental change is accounted for in language learning models (cf. Best's Perception assimilation model: Best 1994, 1995; Flege's speech learning model: Flege 1991; Kuhl's Native Language Magnet: Grieser & Kuhl 1989; Iverson & Kuhl 1996; Kuhl 1991, 1992; Kuhl et al. 1992) as a side effect of the infant having learned the phonological categories of the ambient language during this six to nine month period. It is proposed that by attracting both the ambient and foreign sounds the infant hears, these categories learned by 12 months deafen him to differences detectable six months earlier when he had not yet learned any categories. This defeaning is temporary until apparently 5 to 6 years of age when, if the child does not get sustained exposure to the foreign language early enough, he will be permanently deafened behaviorally to foreign contrasts not shared with the native language. For example, adult speakers of languages with fixed stress (French, Finnish, Hungarian) are significantly less able to detect a shift in stress position in a word than a change in one of its segments (Dupoux, Pallier, Sebastian, & Mehler 1997; Dupoux, Peperkamp, & SebastiánGallés 2001; Peperkamp & Dupoux 2002). Even highly fluent bilinguals are deafened, too, if their exposure to the second language is too late. For example, some Spanish dominant Spanish-Catalan bilinguals who did not learn Catalan before 5–6 years of age cannot discriminate Catalan contrasts not shared with Spanish, high-mid versus low-mid vowels, /e, o/ versus /ɛ, ɔ/, or voiced versus voiceless fricatives, / z / versus / s / (Pallier, Bosch, & SebastiánGallés 1997).

If I am correct in assuming that acoustic differences can always be detected by the bottom-up system, the "deafening", or lack of discrimination mentioned above must be a consequence of the top-down system. The presence or lack of contrasts between sounds is governed by grammatical constraints (Calabrese 2005).[25] Consider the low-back vowel a/ and the front-low vowel /æ/. A language will contrasts these two sounds if and only if the constraint *[+low, −back] is non-active (see Calabrese (2005)). Comparatively, a language has only /a/, and therefore lacks this contrast, if this constraint is active. Constraints are obviously part of the grammar and therefore are also components of the top-down system. In Section 7, I discussed how active constraints trigger repairs that adjust non-native segments in the top-down perceptual system. These repairs lead to perceptual neutralizations of contrasts. When this occurs, two sounds cannot be recognized/identified

---

25. I assume that learning a language involves learning which configurations are admissible. The hypothesis is that the child starts with an inability to produce all segments and combinations of segments except the basic, unmarked ones such as /a/, /m/, /t/, /ta/, /ma/, etc. (see Jakobson (1941)). Learning involves learning to produce the "marked" segments of the ambient language. If the approach proposed here is correct, the child must also be unable to recognize the "marked" sounds before s/he learns to produce them. However, s/he can hear them: s/he has a raw sensation of them in terms of the aural image of the signal present in echoic memory, as proposed earlier. Exposed to the featural configurations extracted from these "raw" stimuli, what I called the sensory intuitions, a child eventually constructs the appropriate combinations of articulatory features in the representations of words and vocabulary items and learns how to articulate them in production and to identify them in perception.

After the end of the critical period, the child loses the ability to easily learn to produce new phonological configurations. A possible way to look at this fact is by noting that the neural motor pathways become set after the critical period so that learning a new array of articulatory movements becomes difficult.

as phonologically/linguistically different. Phonological discrimination means recognition by the top-down systems that two sounds are linguistically/phonologically different. For example, given the case that I discuss in that section, in my own pronunciation of English, there is no contrast between vowels /a/ and /æ/ insofar as I adjust the latter in a different vowel; therefore the contrast between these two vowels is neutralized in my perception of English. I do not discriminate between them in the sense that I cannot recognize them as linguistically different. I can hear that they are acoustically different but I do not know how they are phonologically different, in the sense that I do no know how to account for their difference in articulatory terms, as discussed in Section 7.

Sounds that are recognized as linguistic by the top-down system obviously trigger special linguistic behavioral responses insofar as they are recognized as linguistically relevant in the construction of vocabulary items. Notice that acoustic properties characterizing the foreign sounds will be considered linguistically irrelevant by the linguistic attentional system, and neglected by it—listeners will not pay attention to them, and the acoustic discrepancies brought about by them will be difficult to detect linguistically, although they can be heard granted special attention. Therefore, as Kingston (2003) observes, the deafening observed in children does not mean an inability to hear differences between acoustic stimuli but rather refers to the weakening in the behavioral response to these differences. An infant is behaviorally deafened to foreign contrasts because he no longer responds differently when the stimulus changes from one member of a foreign contrast to the other. Although the infant can still hear the different acoustic stimuli, he just does not respond to them, as it would to native phonological contrast, insofar as they are linguistically unimportant.

For adults, deafening likewise does not imply an inability to hear or access the acoustic signal, but rather a lack of an ability to recognize acoustic configurations as phonological entities, and therefore to discriminate acoustic differences as instances of phonological contrasts. However these acoustic differences can be heard if the adult speaker is made aware of them and pays sufficient attention. If adults were really deaf to foreign sound categories, they would never be able to learn a second language. Indeed, with enough focus adults can hear the phonetic contrasts of a foreign language, and can try to learn how to produce them articulatorily.

This proposal is sufficient cause to warrant a reinterpretation of the Best's findings (1994, 1995). Best investigates how the adaptations of foreign sounds influence the listeners' ability to discriminate different foreign sounds from one another. The primary difference in adaptations is between assimilation of two foreign sounds to two versus just one native sound, "two category" (TC) versus "single category" (SC) assimilations, respectively. Best observes that listeners discriminate the members of TC assimilations far better than SC assimilations. For example, English listeners can discriminate Zulu lateral fricatives /ɬ - ɮ/ well, insofar as they assimilate these two sounds to two different segment categories. However, they do quite poorly with the Thompson Salish ejective velar /k'/ and uvular /q'/ that are likely to assimilate to English [k^h]. SC assimilations are further distinguished between those in which both foreign sounds assimilate equally to the single native category, as in the Zulu lateral fricative case, versus those in which the nonnative pair are both assimilated to a single native category, yet one may be more similar than the other to the native phoneme. This is the case of Zulu aspirated /k^h/ and ejective /kˈ/: they both assimilate to English /k/ but / k^h/ is more similar to the English surface allophone /k^h/. In the latter case, according to Best, the two foreign sounds differ in "category goodness" (CG) with respect to the native category. The members of such CG assimilations are more easily distinguished than the other type of SC. Thus Best suggests that listeners' success in distinguishing different foreign sounds is ranked: TC > CG > SC. She also reports that listeners perform very well in discriminating sounds that cannot be easily assimilated to English sounds like the Zulu clicks.

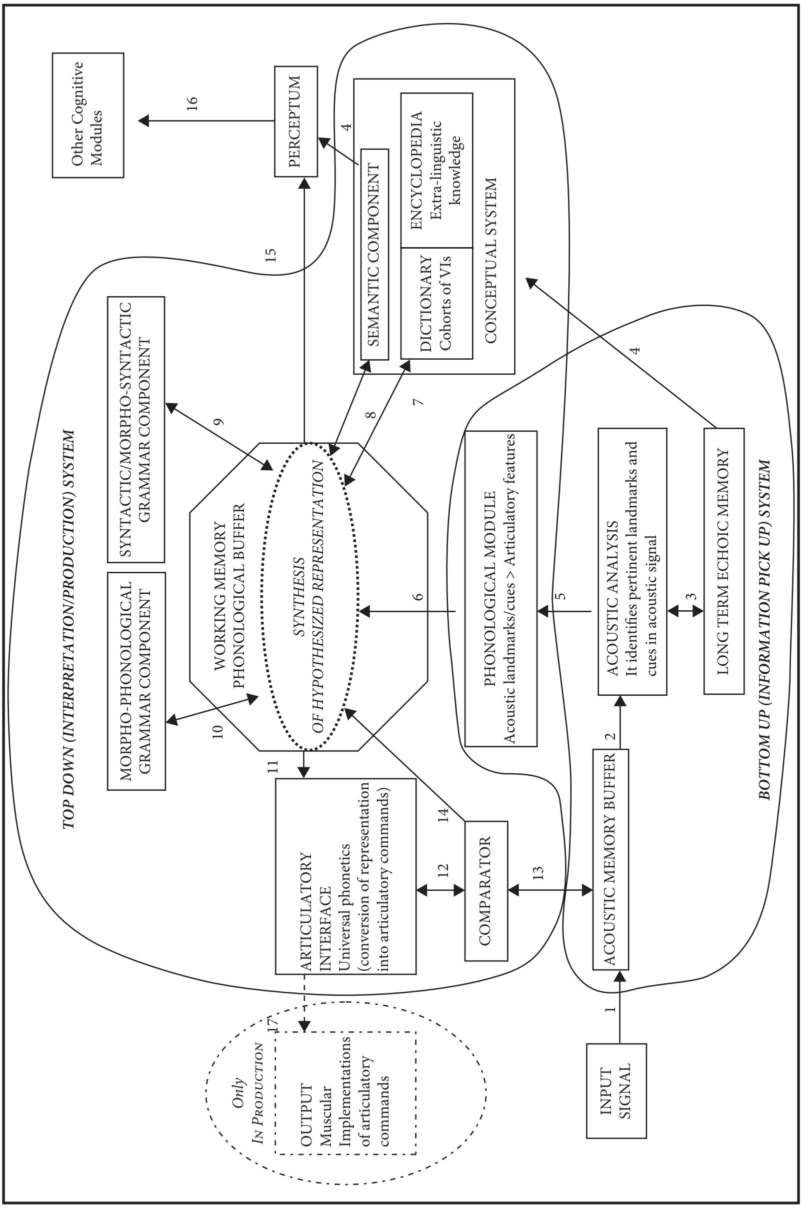
Observe that all of these sounds are acoustically distinct for the listeners regardless of whether they are assimilated to native sounds or not assimilated (as with the clicks). Even in the case of the sounds entering a SC contrast, Best says that they are "heard as discrepant" by listeners (Best 1994:191). What changes is the listeners' capacity to interpret non-native acoustic discrepancies as phonological contrasts by processes of identification in the top-down system. Therefore, if the non-native sounds are identified by the top-down system as involving distinct native phonological categories, they can be discriminated as different. However, if the non-native sounds are identified by top-down interpretation and adjustments as involving a single native phonological category, i.e. if there is perceptual "neutralization", then discrimination is obviously impossible. In the CG, one of the sounds does not undergo any adjustment in the top-down system (as there is no active constraint against it) while the other does. This difference affects the perceptual process, and hence the actual perception of that sound.

10. A model of speech perception

Building on what is outlined above, I propose a speech perception model based on two components: (1) the assumption that both a bottom-up and top-down system are active in perception and (2) the idea that production in the top-down system has a fundamental part in perception of new words/utterances. To do this I turn to the analysis-by-synthesis model of speech recognition proposed by Halle and Stevens (1962) and adopted by Mattingly and Liberman in their motor theory of speech perception.

According to this analysis-by-synthesis model, the listener analyzes the acoustic input by deriving how it is produced by the speaker, synthesizes a virtual acoustic signal based on the output of this derivation, and matches the virtual signal to the actual one. Given a sufficiently close match, the listener achieves a mental representation of the percept that corresponds to the invariant motor commands sent to the musculature underlying the vocal tract actions that produced the acoustic signal. The analysis-by-synthesis component is part of the top-down system. The complete model with the bottom-up and top-down components is schematically represented in (12):

(12)



The block diagram in (12) describes the architecture of the speech perception model proposed here. Italicized numbers in the text refer to the different arrows in (12).[26]

The input acoustic signal enters the bottom-up system and is placed in temporary memory storage pending completion of the analysis *(1)*.

The acoustic analysis in the bottom-up system identifies its invariant acoustic properties and provides a discrete decomposition of the acoustic signal into acoustic landmarks and cues *(2)*. This analysis is generated by a general acoustic module, not specifically dedicated to linguistic signals.

The representation that is so obtained is checked by the long term echoic memory storage system *(3)*. There are two possibilities at this point. (I) If this acoustic representation does not match an already stored representation, that is, it is new or unfamiliar, it needs to be "apprehended", i.e. identified/recognized by the top-down system and goes to step *(5)*. (II) If the acoustic representation matches an already stored representation, the latter becomes active. Once activated, the acoustic representation of familiar words/utterances in the echoic memory system on their turn directly activates the relevant conceptual structures *(4)*. The meaning of an utterance can therefore be directly accessed from the acoustic input in this case. This occurs when we are dealing with commonly used words, constructions and sentences which need to be analyzed only when they are first learned and perceived. Eventually the analysis becomes automatic, and their meaning is automatically associated with the acoustic representation generated in the acoustic analysis phase in the bottom-up system. A new analysis by the top-down system would therefore be unneeded and non-economical. In this case, perception simply bypasses the top-down system.

If the acoustic representation needs to be "apprehended" by the top-down system, it is first sent to the phonological module where the invariant acoustic properties (landmarks and cues) are interpreted as articulatory features *(5)*. For each feature there is a submodule that interprets the acoustic cues/landmarks of the signal and assigns a specification to the feature. Intonational, metrical and other prosodic cues should provide a segmentation of the utterance into phrases and words.

This surface representation of the utterance, which is "hypothetical" in so far as it is the outcome of inferences/interpretation by the phonological module, is sent to the working memory buffer. Here it is parsed/analyzed in the synthesis component in *(12)* and the production system of the working memory buffer *(7)*.

26. I currently consider the architecture in (12) and the relative discussion of this section speculative. It is a way of putting together my hypotheses on how language is produced/perceived and my ideas on how this system of perception/production is positioned in the mind/brain architecture.

A fundamental step in the analysis of this surface representation is the extraction of its underlying representation. This is done by checking cohorts of vocabulary item URs from the Dictionary against the hypothetical surface representation of the words provided by the phonological module. When a vocabulary item UR is chosen for the UR of a word, its morphosyntactic features and meaning is accessed in the dictionary and in the encyclopedia. These morphosyntactic features and meanings are checked against the morphosyntactic and semantic structures that are in the meanwhile generated in the synthesis component to account for the order of the elements and the phrasing in the representation of the utterance *((9)–(10))*. The meaning of the generated structures is also checked against the general pragmatic context. I assume that all of these processes/derivations occur in parallel and will eventually converge in producing a morphosyntactically and semantically well-formed surface sentential representation. Recognition/identification/"apprehension" occur at this point, namely, when a well-formed and licit representation is constructed from the inputs provided by the phonological module. To see if this process is successful, however, it is necessary to wait for the further steps in *(12)–(14)* when a virtual acoustic image of the representation is produced and then compared with the acoustic input stored in the echoic memory buffer, as discussed below.

In fact, the generated surface articulatory representation built from the inputs provided by the phonological module must be checked against the acoustic input to determine whether or not it is correct. This is done by generating a virtual acoustic synthesis of this articulatory representation. First the articulatory representation is converted into complex sets of articulatory commands/gestures *(12)*. These commands are then implemented silently without actual muscular activity, thereby creating a virtual acoustic synthesis that is sent to a comparator module *(13)*. In the comparator, virtual acoustic synthesis is checked against the acoustic input stored in the acoustic memory buffer *(14)*. If there is a successful match, the comparator instructs *(15)* the phonological working memory buffer to read out the representation whose phonological content and morphological, syntactic and semantic structure produced the match and to release it *(16)* as a perceptum to other cognitive modules *(17)*.

At this point it is necessary to consider how the checking and matching of VI URs from the dictionary against the inputs provided by the phonological module is implemented. For the sake of simplicity I consider only what happens in the case of a word simply composed of a root. The same must be done for all of the words and morphemes composing the utterance. There are several possibilities: (I) there is one UR that matches the featural configurations of the word provided by the phonological module; (II) There are more URs that have this perfect match (a case of homonymy); (III) There is a UR that matches the featural configurations of the word provided by the phonological module only partly; or (IV) There are more URs that partly match these featural configurations. Consider possibility I and II first. In case I, there is no problem, the UR is chosen unless the morphosyntactic or semantic context are incompatible with that choice. In case II it is the morphosyntactic and semantic context that determines the selection of the UR. In this case, other similar URs must be tried and the one that is compatible with the context is chosen. In the cases III and IV, it is necessary to access the phonological component *(11)*. The processes (rules and repairs) included in this component are applied to the URs in the relevant order to generate surface representations (see Chapter 14 of Anderson (1992)). The UR of the generated surface representation that matches the acoustic input of the word under analysis is chosen as the UR of this word unless this selection is incompatible with the context. Just as before, if there are more possible selections, the one compatible with the context is chosen. Again I assume that all of these derivations and processes run in parallel until a successful match is reached.

The idea that speech is perceived by reference to production assumes that in the perception process, in particular in what I call *apprehension*, the listener has an active role: he is able to access abstract morphological and syntactic levels of representation of the perceived utterance and compute its surface articulatory shape from these abstract levels. A successful perceptual act occurs when the acoustic shape of the articulatory representation derived in this perceptual computation matches the acoustic input in the auditory memory.

It is important to stress the fundamental role that the production system has in the model. Utterances are generated following the same steps as those discussed in the analysis above. The morphosyntactic component generates the hierarchical organization of the sentence *(10)* which is also computed by the semantic component *(9)*. In this hierarchical structure the UR of the vocabulary items stored in the Dictionary are inserted *(8)*. The surface representation is then derived by applying the phonological and morphophonological processes of the morphophonological component *(11)*. The crucial difference is that the articulatory commands organized in the articulatory interface *(12)* are implemented in the muscular system, and therefore an actual acoustic signal is produced *(18)*.

Observe that in addition to *apprehension* where the listener has an active role, simultaneously a more passive perceptual process can also occur as in the case of commonly used words, constructions and sentences. Although such items are analyzed when they are first learned and perceived, the analysis becomes automatic, and their meaning is automatically associated with the acoustic representation generated in the acoustic analysis phase in the bottom-up system. As proposed above, in this case perception bypasses the phonetic module and working memory.

The model in (12) predicts both Top-down and Bottom-up effects in speech perception. These kinds of findings are often described as evidence for an interaction of "bottom-up" and "top-down" processes in perception (e.g. Klatt. 1980). Bottom-up processes analyze the acoustic signal as it comes in. Top-down processes draw inferences concerning the signal based both on the fragmentary results of the continuing bottom-up processes and on stored knowledge of likely inputs. As discussed in the introduction, top-down processes can restore missing phonemes or correct erroneous ones in real words by comparing results of bottom-up processes against lexical entries, or they can generate gross departures of perceptual experience from the stimulus as observed in mishearings.

The model in (12) accounts both for bottom-up analytic processes and a top-down constructions and restoration processes. They interact in the phonological working memory buffer where the structures underlying heard utterances are constructed.

Consider how new words are learned according to (12). First of all, in this case, the dictionary will not play any role in the analysis. The crucial component is the phonetic module that provides the featural inputs of the new words to the synthesis component that then constructs their complete representation. In particular it hypothesizes possible underlying representation for the new word and then derives their surface representation by applying to it the processes (rules/repairs) of L1. In the case of foreign words with unfamiliar sounds, the featural input provided by the phonological module cannot be used to construct licit featural representations of segments/syllabic configurations. The synthesis component must then adjust the featural input and construct representations that are licit according to the grammar of L1.

11. The construction of underlying representations

As proposed above, learning a word requires an analytical process that involves the grammatical knowledge that is used in production. In this section, I consider the analytical process involved in the construction of URs and demonstrate that the URs that are constructed in the case of foreign words must be consistent with L1 grammar. They must be "familiar", or "interpretable" in terms of grammar of L1,.

Before committing a foreign word to long-term memory, its underlying representation must be constructed. In Generative Grammar it is assumed that a UR ABC is postulated for a surface form AED in a language L when phonological alternations or distributional patterns in L provide evidence for two ordered processes in (13):

(13) a. C→ D/ ___ #
b. B→ E/ ___ D

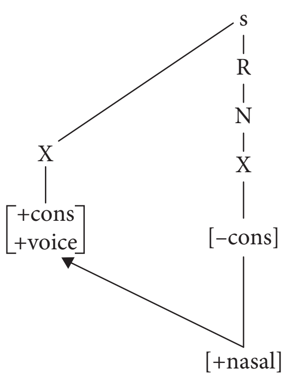
As is well known, postulating a base form implies postulating a rule, and viceversa, given the evidence provided by the alternations in the language. The L2 learner may have a limited access to the alternations needed to identify the UR of the L2 words. This limited access may be due to the time constraints of language acquisition or to the fact that as an adult the learner no longer has the ability to recognize and appropriately analyze the phonological alternations of L2, Therefore faced with a non-native form, the learner tends to analyze it in terms of his L1 system.

Given the analysis-by-synthesis model proposed here, the analysis of a word requires the reverse application of the phonological derivation. Therefore given the two rules in *(13)*, if the surface shape of the word is AED, an underlying ABC must be postulated so as to derive the surface AED.

In many cases the postulated UR does not need to be different from the surface L2 form. For example, if the L1 has an underlying distinction between voice vs. voiceless obstruents, and word-final devoicing, as in German, a speaker could assume an UR /gUd/ for English "good", despite the fact that he will pronounce this word [gUt].

In some cases, however, a different UR must be postulated. Take the Brazilian indigenous language Maxacalí. Wetzels (this volume) shows that Maxacalí nasality is contrastive only in the case of vowels. Nasal consonants are always derived by spreading the nasal feature of this vowel. In particular there is a rule spreading nasality from the vowel onto the syllabic onset of this vowel, making words such as *[bãŋõn] impossible in this language.

(14)



Wetzels shows that in Brazilian Portuguese (BP) loanwords to Maxacalí, the original nasal onsets of the loanwords are analyzed as being the outcome of this spreading rule. As he puts it, "confronted with a nasal onset of an oral syllable, the speaker of Maxacalí interprets the nasal onset as an indication of the nasality of its nucleus." Therefore faced with BP words such as those in (15), a Maxacalí speaker postulates a UR where the nasality is a property of the vowel. The rule in (14) then spreads the nasality onto the preceding voiced stop onset. The result is that the speaker postulates a UR consistent with the L1 phonological system.

(15)			BP	Maxacalí
	Margarida	'Margarida'	[mahgaɾida]	[mã^{ŋ}gadit]
	carneiro	'sheep'	[kahnejɾω]	[kahnɛ̃n]
	mesa	'table'	[mezɐ]	[mẽ^{n}d^{ʒ}a]
	moto	'motorbike'	[mɔtω]	[mõtok]

Observe that if the vowel is interpreted as [−nasal] in the borrowing, its onset is also non- nasal, just as expected if nasality is a property of the vowel and nasality in onset is derived by rule.

(16)			BP	Maxacalí
	martelo	'hammer'	[mahtɛlω]	[^{m}bahtɛt]
	canivete	'pocketknife'	[kanivɛt^{ʃ}i]	[kɯdibɛt]

Awareness of the rules and constraints of the L1 grammar, therefore, leads to the postulation of more abstract representations for L2, in particular the postulation of a representation for L2 consistent with the rules and constraints of L1. Consider some other examples. Nevins and Braun (this volume) discuss the following case involving the pronunciation of English by Brazilian Portuguese speaker in light of a BP rule changing the rhotic /r/ to a laryngeal fricative in word-initial position:

(17) direto [dʒirɛtu] vs. reto [hetu]

Interestingly, in their pronunciation of English, Brazilian speakers replace word-initial /h/ with [r]:

(18)		BP pronunciation	
	home	[rom]	(or [hom])
	hug	[rʌg]	(or [hʌg])
	hunger	[rʌ̃gər]	(or [hʌ̃gər])

Nevins and Braun propose that when exposed to English words, a Brazilian learner observes that the rule debuccalizing [r] into [h] does not apply to English. When faced to word-initial /h/ in English, he then hypothesizes that it derives from underlying /r/ as in his own language. Given that he has postulated that r-debuccalization does not apply in English, this hypothesized /r/ surfaces in the English word as can be seen in (18). Again the UR postulated for these words is consistent with the grammar of L1.

Another example involves my own pronunciation of English. Italian does not have the laryngeal fricative /h/. When I speak English, I delete it especially when word initial. However, I also often insert a laryngeal fricative in the same context.

(19)

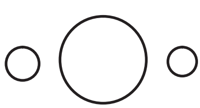
	My pronunciation
harbor	[arbor]
aisle	[hayl]

A possible analysis would be that in the UR of the relevant English words in my long term memory, I do not have laryngeal fricatives as required by the Italian grammar. Given that I observe that English has /h/, especially in word-initial position, I hypothesize a rule of /h/-insertion to mask my problem. The point is the UR of English h-initial words in my own lexicon is consistent with my own native grammar.

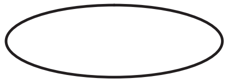
12. Galileo, Saturn and the pharyngeal vowels

In this concluding section, I will consider Jacobs and Gussenhoven's (2000) objection to the idea that a model that assumes that foreign sounds are modified in perception predicts that people cannot hear segmental contrasts that do not occur in their own language. According to this argument, if this is true, we would expect major problems when the speaker of languages with small segment inventories, like Tahitian or Maori are exposted to languages with larger inventories such as English. For example, as the former languages lack the /t/ and /s/, one may predict that speakers of these languages would be incapable of hearing the difference between these two segments given that their languages do not have this opposition. However, as Jacobs and Gussenhoven observe, this is difficult to reconcile with the common finding that language users appear to be capable of hearing (at least some) non-native segments with ease. I agree with them. Kant also observed that raw sensation is different from interpretation; i.e. having a sensation is different from understanding it. A monolingual Tahitian or Maori speaker will hear that the English /s/ and /t/ are different sounds. The issue is that he cannot understand how they are produced, and therefore their identification remains fuzzy and unclear.

With this in mind, I want to draw a parallel from the history of science to explain this idea better. Eco (1977) conveys that when Galileo looked at Saturn for the first time, he saw something never seen before. In his various letters to friends and colleagues, Galileo described the efforts he made (as he looked) to understand the shape of Saturn. For example, in three letters (to Benedetto Castelli, 1610, to Belisario Giunti, 1610; and to Giuliano de Medici, 1611), he says he saw not one star but three joined together in a straight line parallel to the equinoctial; he represented this in an drawing like the one below:

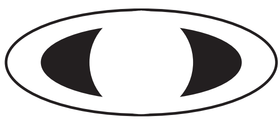
(20) 

But in other letters (e.g. to Giuliano de Medici, 1610; and to Marco Velseri, 1612) he admits that owing "to the imperfection of the instrument and the eye of the observer", Saturn might also appear, as in (21), "in the shape of an olive".

(21) 

The figure clearly reveals that, since it is wholly unexpected for a planet to be surrounded by a ring (which apart from anything else clashed with every notion held at the time with regards to heavenly bodies), Galileo was trying to understand what he could see; he was laboriously attempting to construct a (new) mental representation of Saturn.

After looking at the star and studying the situation for some time, (see his letter to Federigo Borromeo in 1610) Galileo finally decided that it was not a matter of two small round bodies but of larger bodies "and of a shape no longer round, but as can be seen in the enclosed figure, two semi-ellipses with two very obscure little triangles in the middle of the said figures, and contiguous to Saturn's middle globe." This consideration led Galileo to a third representation, (22):

(22) 

Note that Galileo did not recognize the existence of rings, otherwise he would have written not of two semi-ellipses but of an elliptical band. It is only in trying to convey on paper the essential features of what he observed that Galileo gradually began to "see", to perceive Saturn and its rings. He finally "understood" its nature. Prior to that, Galileo could not recognize or identify what he was seeing, and he had to interpret it by trying different mental representations.

Observe that the sensation, or better the "sensory intuition", is still there before the interpretation. Our interpretation of the world does not change that. Therefore, Galileo was obviously able to distinguish what he was seeing in the case of Saturn from what he was seeing in the case of Jupiter. If, after his first viewing of Saturn, he had been asked if this new planet resembled Jupiter, he would have answered that it did not.

I claim that the same occurs in our perception and representation of foreign sounds. Once I attended a field method course with an Abhkaz speaker. In one of the classes, the informant uttered words with pharyngealized vowels. I had the distinct feeling that they were different from the plain ones, but I was not certain about the nature of this difference. Upon a second hearing, I mistakenly perceived them as fronted vowels, as diphthongs composed of a plain vowel plus something else, and even as plain vowels. Only in later classes, after being told how to pronounce them, and having read some literature on the topic and having practiced pronouncing them, did I gain the ability to distinguish them from the plain ones in words that were pronounced slowly. Only after all this background could I begin to identify them, albeit only tentatively and in slow speech.

I wonder how to characterize this learning that had occurred to me. I heard these strange sounds, which were totally new to me (though admittedly there is a certain advantage in being a phonologist and knowing that pharyngealized vowels exist). While hearing utterances with these vowels, my phonological module constructed a first articulatory mental representation of these words, a sensory intuition. In the case of the pharyngealized vowels, the representations may have included the feature [+RTR]. In the phonological working memory buffer where phonology is accessed, incomplete representations or representation with the feature combination [−consonantal, +RTR] were illicit, and had to be adjusted by phonological operations, i.e. by repairs, in an attempt to produce a recognizable, familiar representation of these sounds, so as to apprehend them. These were illusory representations.

When, with training and effort, I learned to coordinate a [+RTR] configuration of the tongue root with a [−consonantal] stricture, I was able to match my internal representation of these sounds with their acoustic shape and I recognized them more or less well. But this was just tentative, and temporary, insofar as being an adult, I could not learn to articulate pharyngeal vowels, and I will always have both articulatory and perceptual problems with them. Obviously all of this was occurring in a very special context, a field method course, and I was being taught about pharyngeal vowels. Normal speakers are not so lucky and they will normally stop at the stage of the illusory representations. They may indeed feel that foreign sounds are auditorily different from other sounds, but they cannot identify or understand them because they cannot articulate them. In this case they will adjust them phonologically into sounds that are licit and articulatorily possible.

Articulating a sound and perceiving it, in the sense of apprehending it, is the same thing. To conclude with Giovann Battista Vico (1688–1744), *verum et factum reciprocantur seu convertuntur*. The human mind can know only what the human mind has made. Still, reality (acoustic reality, in the cases discussed here) is out there to check us, to control us, to stimulate us to change…

References

Aboitiz, Francisco & Ricardo V. Garcia. 1997. "The evolutionary origin of language areas in the human brain. A neuroanatomical perspective". *Brain Research Reviews* 25. 381–396.

Aslin, Richard N., Peter W. Jusczyk, & David B. Pisoni. 1998. "Speech and auditory processing during infancy: Constraints on and precursors to language". *Handbook of child psychology:*

Cognition, perception, and language ed. by Deanna Kuhn & Robert Siegler, 147–254. New York: Wiley.

Anderson, John. 1985. *Cognitive Psychology and Its Implications*. New York: Freeman.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Awh Edward, John Jonides, Edward E. Smith, Eric H. Schumacher, Robert A. Koeppe & Stewart Katz. 1996. "Dissociation of storage and rehearsal in working memory: PET evidence". *Psychological Science* 7. 25–31.

Baddeley, Alan D. 1992. "Working memory". *Science* 255. 556–559.

Best, Catherine T. 1994. "The emergence of native language phonological influences in infants: A perceptual assimilation hypothesis". *The development of speech perception* ed. by Judith C. Goodman & Howard C. Nusbaum.167–224. Cambridge, MA: MIT Press.

Best, Catherine T. 1995. "A direct realist perspective on cross-language speech perception". *Crosslanguage speech perception* ed. by Winifred Strange & James J. Jenkins, 171–204. Timonium, MD: York Press.

Best, Catherine T., Gerald W. McRoberts, Rosemarie Lafleur, & Jean Silver-Isenstadt. 1995. "Divergent developmental patterns for infants' perception of two nonnative speech contrasts". *Infant Behavior and Development* 18. 339–350.

Blumstein Sheila & Kenneth N. Stevens. 1981. "Phonetic features and acoustic invariance in speech". *Cognition* 10. 25–32.

Boberg, Charles. 1999. "The Attitudinal Component of Variation in American English Foreign a". *Nativization, Journal of Language and Social Psychology* 18.1. 49–61.

Bobaljik, Jonathan D. 2006. "Itelmen Reduplication : Edge-in Association and Lexical Stratification". *Journal of Linguistics* 42. 1–23.

Boersma, Paul & Silke Hamann. This Volume. "Loanword adaptation as first-language phonological perception ".

Boring, Edwin G. 1946. "Perception of objects". *American Journal of Physics* 14. 99–107.

Bromberger, Sylvain & Morris Halle. 1992. "The ontology of Phonology". *What we know we don't know* by Sylvain Bromberger, 209–228. Chicago: Chicago University Press.

Bromberger, Sylvain & Morris Halle. 1997. "The content of phonological signs: a comparison between their use in Derivational Theories and in Optimality Theories". *Derivations and Constraints in Phonology* ed. by Iggy Roca, 93–122. Oxford: Oxford University Press.

Bromberger, Sylvain & Morris Halle. 2000. "The ontology of Phonology revised." *Phonological Knowledge. Conceptual and Empirical Issues* ed. by Noel Burton-Roberts, Philip Carr, & Gerald Docherty, 19–37. Oxford: Oxford University Press.

Browman, Catherine. 1980. "Perceptual processing: evidence from slips of the ear". *Errors in linguistic Performance: Slips of the tongue, ear, pen, and hand* ed. by Victoria Fromkin. New York: Academic Press.

Calabrese, Andrea. 1988. *Towards a theory of phonological alphabets*. Ph.D. diss., Massachusetts Institute of Technology.

Calabrese, Andrea. 1995. "A constraint-based theory of phonological markedness and simplification procedures". *Linguistic Inquiry* 26.2. 373–463.

Calabrese, Andrea. 2005. *Markedness and Economy in a Derivational Model of Phonology,* Mouton-De Gruyter, Berlin.

Chistovich, Ludmilla A. & Valentina V. Lublinskaya. 1979. "The 'center of gravity' effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli". *Hearing Research* 1. 185–195.

Dehaene-Lambertz, Ghislaine. 1997. "Electrophysiological correlates of categorical phoneme perception in adults". *NeuroReport* 84.3. 919–924

Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux & A. Gout. 2000. "Electrophysiological correlates of phonological processing: a cross-linguistic study". *Journal of cognitive neuroscience* 12. 635–647.

Delgutte, Bertrand, & Nelson Y. S. Kiang. 1984. "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics," *Journal of the Acoustic Society of America.* 75, 897–907.

Diehl, Randy L., Andrew J. Lotto, & Lori L. Holt. 2004. "Speech perception". *Annual Review of Psychology* 55. 149–179.

Di Pellegrino, Giuseppe, Luciano Fadiga, Leonardo Fogassi, Vittorio Gallese, & Giacomo Rizzolatti. 1992. "Understanding motor events: A neurophysiological study". *Experimental Brain Research* 91. 176–180.

Doupe, Allison J., & Patricia K. Kuhl. 1999. "Birdsong and human speech: common themes and mechanisms". *Annual Review of Neuroscience* 22. 567–631.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. "Epenthetic vowels in Japanese: A perceptual illusion?". *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian-Galles & Jacques Mehler. 1997. "A destressing "deafness" in French?". *Journal of Memory and Language* 36. 406–421.

Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi & Jacques Mehler. 2001. "New evidence for prelexical phonological processing in word recognition". *Language and Cognitive Processes* 16. 491–505.

Dupoux, Emmanuel, Peperkamp, Sharon, & Nuria Sebastian-Galles (2001). "A robust method to study stress 'deafness'". *Journal of the Acoustical Society of America*, 110, 1606–1618.

Eco, Umberto. 1997. *Kant and the Platypus. Essays on Language and Cognition*. New York: Harcourt Brace.

Epstein, William. 1982. "Percept-percept couplings". *Perception* 111. 75–83

Fadiga, Luciano, Laila Craighero, Giovanni Buccino, & Giacomo Rizzolati. 2002. "Speech listening specifically modulates the excitability of tongue muscles: A TMS study". *European Journal of Neuroscience* 15. 399–402.

Fitzgerald, B.H. 1996. *Acoustic Analysis of Japanese Vowels Produced in Multiple Consonantal Contexts*. MA thesis, University of South Florida.

Flege, Jim E. 1991. "Perception and production: The relevance of phonetic input to L2 phonological learning." *Crosscurrents in second language acquisition and linguistic theories* ed. by Charles Ferguson & Thom Huebner. Philadelphia, PA: John Benjamins.

Fowler, Carol A. 1986. "An event approach to the study of speech perception from a direct-realist perspective". *Journal of Phonetics* 14. 3–28.

Fowler, Carol A. 1994. "Invariants, specifiers cues: An investigation of locus equations as information for place of articulation". *Perception & Psychophysics* 55. 597–610.

Fowler, Carol A. 1996. "Listeners do hear sounds, not tongues". *Journal of the Acoustical Society of America* 99. 1730–1741.

Fowler, Carol A. & Bruno Galantucci. 2005. "The relation of speech perception and production". *The handbook of speech perception* ed. by David B. Pisoni & Robert E. Remez. 633–652. Malden, MA: Blackwell.

Fowler, Carol A., Philip Rubin, Robert Remez, & Michael Turvey. 1980. "Implications for speech production of a general theory of action". *Language production: Vol.* 1. *Speech and talk* ed. by Brian Butterworth, 373–420. London: Academic Press.

Fox, Robert. 1984. "Effects of lexical status on phonetic categorization". *Journal of Experimental Psychology: Human Perception and Performance* 10. 526–540.

Galantucci, Bruno, Carol A. Fowler, & Michael T. Turvey. 2006. "The Motor Theory of Speech Perception Reviewed". *Psychonomic Bulletin and Review* 13.3. 361–377.

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

Garnes, Sara & Zinny Bond. 1980. "A slip of the ear: A snip of the ear? A slip of the year?". *Errors in Linguistic Performance: Slips of the tongue, ear, pen, and hand* ed. by Victoria Fromkin. New York: Academic Press.

Gibson, James J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gregory, Richard L. 1993. "Seeing & thinking". *Giornale Italiano di Psicologia* 205. 749–769

Grieser DiAnne & Patricia Kuhl. 1989. "Categorization of Speech by Infants: Support for Speech-Sound Prototypes". *Developmental Psychology* 25.4. 577–88

Goldstone Robert L. & Alan Kersten. 2003. "Concepts and Categorization". *Comprehensive Handbook of Psychology* ed. by Alice Healy & Robert Proctor, 599–621. New Jersey: Wiley

Goldinger, Stephen D. 1996. "Words and voices: Episodic traces in spoken word identification and recognition memory". *Journal of Experimental Psychology: Learning, Memory and Cognition* 22. 1166–1183.

Goldinger, Stephen D. 1998. "Echoes of echoes? An episodic theory of lexical access". *Psychological Review* 105. *251–279.*

Goldinger, Stephen D. 2000. "The role of perceptual episodes in lexical processing". *Proceedings of the Workshop on Spoken Word Access Processes SWAP* ed. by Ann Cutler, J.M. McQueen, & R. Zondervan, 155–159. Nijmegen: Max Plack Institute for Psycholinguistics.

Guion, Susan G. 2003. "The vowel systems of Quichua Spanish bilinguals: An investigation into age of acquisition effects on the mutual influence of the first and second languages". *Phonetica* 60. 98 -128.

Hale Mark & Charles Reiss. 1998. "Formal and Empirical Arguments concerning Phonological Acquisition". *Linguistic Inquiry* 29. 656–683 .

Halle, Morris. 1983. "On distinctive features and their articulatory implementation". *Natural Language and Linguistic Theory* 1.1. 91–105

Halle, Morris. 1995. "Feature geometry and feature spreading". *Linguistic Inquiry* 26. 1–46.

Halle, Morris. 2002. *From memory to speech and back. Papers on Phonetics and Phonology* 1954–2002. Berlin: Mouton De Gruyter.

Halle, Morris & Kenneth Stevens. 1962. "Speech recognition: A model and a program for research". *IEEE Transactions on Information Theory* 8. 155–159. Now in Halle 2002. 25–36.

Halle, Morris & Kenneth Stevens. 1991. "Knowledge of Language and the Sounds of Speech". *Music. Language, Speech and Brain* ed. by J. Sundberg et alii, 1–91. McMillan Press, London. Now in Halle 2002. 176–195.

Halle, Morris, Bert Vaux & Andrew Wolfe. 2000. "On feature spreading and the Representation of Place of Articulation". *Linguistic Inquiry* 31. 387–444.

Herd, Jonathan. 2005. "Loanword adaptation and the evaluation of similarity". *Toronto Working Papers in Linguistics* 24: *Special issue on similarity in phonology* ed. by Chiara Frigeni, Manami Hirayama & Sara Mackenzie, 65–116.

Helmholtz, Herman Von. 1867. Handbuch der Physiologischen Optik. Hamburg und Leipzig : Leopold Voss

Hickock, Gregory & David Poeppel. 2004. "Dorsal and Ventral Streams: a framework for understanding aspects of the functional anatomy of language". *Cognition* 92. 67–99.

Hintzman, Douglas 1986. "Schema abstraction in a multiple trace memory model". *Psychological Review* 93.411–428.
Homma, Yayoi 1981. "Durational relationship between Japanese stops and vowels". *Journal of Phonetics* 9. 273–281.
Houde, John F. & Michael I. Jordan. 1998. "Sensorimotor Adaptation in Speech Production." *Science* 20. 279. 1213–1216
Hyman, Larry 1970. "The role of borrowings in the justification of phonological grammars." *Studies in African linguistics* 1. 1–48.
Iacoboni, Marco, Roger P. Woods, Marcel Brass, Harold Bekkering, John C. Mazziotta & Giacomo Rizzolatti.1999. "Cortical mechanisms of human imitation". *Science* 286. 2526–2528.
Ingram, John & Park See-Gyoon. 1998. "Language, context, and speaker effects in the identification and discrimination of English */r/* and */l/* by Japanese and Korean listeners". *Journal of the Acoustical Society of America* 103. 1161–1174.
Ito, Chiyuki, Yoonjung Kang, & Michael Kenstowicz. 2006. "The Adaptation of Japanese Loanwords into Korean". *MIT Working Papers* 52. 65–104.
Iverson, Gregory & Ahrong Lee. 2004. "Perceived syllable structure in the adaptation of loanwords in Korean". Ms., University of Wisconsin, Milwaukee. [Presented at Workshop on Loanword Adaptation, 2004 Linguistic Society of Korea International Conference, Seoul, July 28–30]
Iverson, Paul, & Patricia K. Kuhl. 1995. "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling". *Journal of the Acoustical Society of America* 97. 553–562.
Iverson, Paul, & Patricia K. Kuhl. 2000. "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?". *Perception and Psychophysics* 62. 874–886.
Jacobs, Haike & Carlos Gussenhoven. 2000. "Loan phonology: perception, salience, the lexicon and OT". *Optimality Theory: Phonology, syntax, and acquisition* ed. by Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer, 193–210. Oxford: Oxford University Press.
Jakobson, Roman. 1941. *Kindersprache, Aphasie, und allgemeine Lautgesetze* [Child Language, Aphasia and Phonological Universal]. Uppsala Universitets Aarskrift. Translated into English by A. Keiler. The Hague: Mouton, 1968.
Kabak, Baris & William Idsardi 2006. "Speech Perception is not isomorphic to phonology". ms. University of Konstanz, Konstanz.
Kanisza, Gaetano 1979. *The Organization in Vision: Essays on Gestalt Perception*, New York, Praeger.
Kenstowicz, Michael & Hyang-Sook Sohn. 2001. "Accentual adaptation in North Kyungsang Korean". *Ken Hale. A Life in Language* ed. by Michael Kenstowicz, 239–270. Cambridge, MA: MIT Press.
Kenstowicz , Michael. 1994. *Phonology in Generative Grammar*. Blackwell, Cambridge.
Kenstowicz, Michael. 2001. "The role of perception in loanword phonology". *Studies in African Linguistics* 32. 95–111.
Kenstowicz, Michael. 2003. "Salience and similarity in loanword adaptation: a case study from Fijian". *Language Sciences* 29. 316–340.
Kingston, John. 2003. "Learning Foreign Vowels language and speech". *Language and Speech* 46. 295–349.
Kim, Hyunsoon. This volume. "Korean adaptation of English affricates and fricatives in a feature-driven model of loanword adaptation".

Klatt, Dennis. 1980. "Speech perception: A model of acoustic-phonetic analysis and lexical access". *Perception and production of fluent speech* ed. by Ronald Cole. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kohler Evelyne, Christian Keysers, M. Alessandra Umiltà, Leonardo Fogassi, Vittorio Gallese & Giacomo Rizzolatti. 2002. "Hearing sounds, understanding actions: Action representation in mirror neurons". *Science* 297. 846–848.

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

Kuhl, Patricia K. 1993. "Innate predispositions and the effects of experience in speech perception: The native language magnet theory". *Developmental neurocognition: Speech and face processing in the first year of life* ed. by Benedicte de Boysson Bardies, Scania de Schonen, Peter Jusczyk, McNeilage, Peter & John Morton, 259–274. Dordrecht, Netherlands: Kluwer Academic Publishers.

Kuhl, Patricia K. 2000. "Language, mind, and brain: Experience alters perception." *The new cognitive neurosciences* ed. by Michael S. Gazzaniga, 99–115. Cambridge, MA: MIT Press.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens, & Björn Lindblom. 1992. "Linguistic experience alters phonetic perception in infants by six months of age". *Science* 255. 606–608.

LaCharite, Darlene & Carole Paradis. 2005. "Category preservation and proximity vs. phonetic approximation in loanword adaptation". *Linguistic Inquiry* 36. 223–58.

Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell Publishers.

Liberman, Alvin M. 1957. "Some results of research on speech perception". *Journal of the Acoustical Society of America* 29. 117–123.

Liberman, Alvin M. 1996. *Speech: A special code.* Cambridge, MA: Bradford Books.

Liberman, Alvin M., Franklin S. Cooper, Donald Shankweiler & Michael Studdert-Kennedy. 1967. "Perception of the speech code". *Psychological Review* 74. 431–461.

Liberman, Alvin M., Pierre Delattre & Franklin S. Cooper. 1952. The role of selected stimulus variables in the perception of the unvoiced-stop consonants. *American Journal of Psychology,* 65. 497–516.

Liberman. Alvin M. & Ignatius G. Mattingly. 1985. "The motor theory or speech perception revised". *Cognition* 21. 1–36.

Liberman. Alvin M. & Ignatius G. Mattingly. 1989. "A specialization for speech perception". *Science* 243. 489–494.

Liberman, Alvin M., & Doug H. Whalen. 2000. "On the relation of speech to language". *Trends in Cognitive Sciences* 4. 187–196.

Mann, Virginia & Alvin Liberman. 1983. "Some differences between phonetic and auditory modes of perception". *Cognition* 14. 211–23.

Marr , David. 1982. *Vision*. San Francisco: Freeman.

Marslen-Wilson, William & Alan Welsh. 1978. "Processing interactions and lexical access during word recognition in continuous speech". *Cognitive Psychology* 10. 29–63.

McCawley, James 1968. *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

McClelland, James & Jeff Elman. 1986. "The TRACE model of speech perception". *Cognitive Psychology* 18.1–86.

Merleau-Ponty, Maurice. 1945. *Phénoménologie de la perception*, Paris: Gallimard, 1945.Trans. by Colin Smith, (New York: Humanities Press, 1962.

Milner, A. David & Melvyn A. Goodale 1995. *The visual brain in action*. Oxford: Oxford University Press.

Näätänen, Risto. 2001. "The perception of speech sounds by the human brain as reflected by the mismatch negativity MMN and its magnetic equivalent MMNm". *Psychophysiology* 38. 1–21.

Neisser, Ulric. 1967. *Cognitive psychology*. Appleton-Century-Crofts New York.

Nevins, Andrew & David Braun. This Volume. "The role of underlying representations in L2 Brazilian English".

Norman, Joel. 2001. "Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches". *Behavioral and Brain Sciences* 25. 73–96.

Okada, Hideo 1999. "Japanese". *Handbook of the International Phonetic Association:* A *guide to the usage of the International Phonetic Alphabet.* 117–119 Cambridge: Cambridge University Press.

Paavilainen, Petri, M. Jaramillo, Risto Näätänen, & Istvan Winkler. 1999. "Neuronal populations in the human brain extracting invariant relationships from acoustic variance". *Neuroscience Letters* 265. 179–182.

Paavilainen, Petri, J. Simola, M. Jaramillo, Risto Näätänen & Istvan Winkler. 2001. "Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity MMN". *Psychophysiology* 38. 359–365.

Pallier, Christophe, Laura Bosch, Nuria Sebastian Gallés. 1997. "A limit on behavioral plasticity in speech perception". *Cognition* 643. 9–17.

Paradis, Carole. 1992. *Lexical Phonology and Morphology: The Nominal Classes in A Psycholinguistic Theory of Loanword Adaptations in Fula.* New York: Garland.

Paradis, Carole. 1995. "Native and loanword phonology as one: constraints versus rules". *Proceedings of the 13th International Congress of Phonetic Sciences*, ed. by K. Elenius & P. Branderud, 374–81. Stockholm.

Paradis, Carole & Antoine Tremblay. This Volume. "Nondistinctive Features in Loanword Adaptation: The Unimportance of English Aspiration in Mandarin Chinese Phoneme Categorization."

Peperkamp, Sharon. 2002. "Towards a new theory of loanword adaptations". Ms., University of Paris 8.

Peperkamp, Sharon, & Emmanuel Dupoux. 2002. "A typological study of stress "deafness"." *Laboratory phonology 7* ed.by Carlos Gussenhoven & Natasha Warner, 203–240. Berlin: Mouton de Gruyter.

Peperkamp, Sharon & Emmanuel Dupoux. 2003. "Reinterpreting loanword adaptations: the role of perception". *Proceedings of the I5th International Congress of Phonetic Sciences.* 367–70.

Phillips, Colin. 2001. "Levels of representation in the electrophysiology of speech perception". *Cognitive Science 25.* 711–731.

Phillips, Colin, Thomas Pellathy, Alec Marantz, Elron Yellin, Ken Wexler, David Poeppel, Martha McGinnis, Timothy Roberts. 2000. "Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study". *Journal of Cognitive Neuroscience* 12. 1038–1055.

Piaget Jean. 1937. *La construction du reel chez l'enfant*. Neuchatel: Delachaux et Niestle

Piaget, Jean. 1969. *Mechanisms of perception* New York: Basic Books.

Polka, Linda, & Janet F. Werker. 1994. "Developmental changes in perception of nonnative vowel contrasts". *Journal of Experimental Psychology, Human Perception, and Performance* 20. 421–435.

Pulvermüller, Friedemann, Martina Huss, Ferath Kherif, Fermin Moscoso Del Prado Martin, Olaf Hauk & Yuri Shtyrov. 2006. "Motor cortex maps articulatory features of speech sounds". *Proceedings of the National Academy of Sciences* 103. 7865–7870.

Pulvermüller, Friedemann, Teja Kujala, Yuri Shtyrov, Jaana Simola, Hannu Tiitinen, Paavo Alku, Kimmo Alho, S. Martinkauppi, Risto J. Ilmoniemi, & Risto Näätänen. 2001. "Memory traces for words as revealed by the Mismatch Negativity MMN". *NeuroImage* 14. 107–616.

Repetti, Lori. 2003. "Come i sostantivi inglesi diventano italiani: la morfologia e la fonologia dei prestiti". *Italiano ed Inglese a confronto*, ed. by Sullam Calimani, 31–42. Firenze: Giunti.

Repetti, Lori. 2006. "The emergence of Marked Structures in the Integration of Loans in Italian" *Historical Romance Linguistics. Retrospectives and Perspectives*, ed. by Randall S. Gess & Deborah Arteaga. Amsterdam John Benjamins.

Repetti, Lori. This volume. "Gemination in English loans in American varieties of Italian"

Rizzolatti, Giacomo & Michael A. Arbib. 1998. "Language within our grasp". *Trends in Neurosciences* 21. 188–194.

Rizzolatti, Giacomo, R. Camarda, Leonardo Fogassi, M. Gentilucci, G. Luppino, & Massimo Matelli. 1988. "Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements". *Experimental Brain Research* 71. 491–507.

Rizzolatti, Giacomo, & Laila Craighero. 2004. "The mirror-neuron system". *Annual Review of Neuroscience* 27. 169–192.

Rizzolatti, Giacomo, Leonardo Fogassi, & Vittorio Gallese. 2001. "Neurophysiological mechanisms underlying the understanding and imitation of action". *Nature Reviews Neuroscience* 2. 661–670.

Rock, Irving. 1977. "In defense of unconscious inference." *Stability and constancy in visual perception: Mechanisms and processes* ed. by William Epstein, 321–377. New York.

Rock, Irving. 1983. *The logic of perception*. Cambridge, MA: MIT Press.

Rock, Irving. 1997. *Indirect perception*. Cambridge, MA: MIT Press.

Samuel Arthur G. 1981. "Phonemic Restoration: Insights from a New Methodology." *Journal of Experimental Psychology* 110. 474–494.

Sapir, Edward. 1933. "The Psychological Reality of Phonemes". *Selected Writings of Edward Sapir in Language, Culture and Personality* ed. by David G. Mandelbaum, 46–60. Berkeley: University of California Press. 1949.

Shestakova, Anna; Elvira Brattico, Minna Huotilainen, Valery Galunov, Alexei Soloviev, Mikko Sams, Risto J. Ilmoniemi, Risto Näätänen. 2002. "Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study". *Neuroreport* 1314. 1813–1816.

Shinohara, Shikego. 1997. *Analyse phonologique de l'adaptation japonaise de mots etrangers.* Doctoral dissertation, Universite de Paris III.

Shinohara, Shikego. 2000. "Default accentuation and foot structure in Japanese: evidence from Japanese adaptations of French words". *Journal of East Asian Linguistics* 9. 55–96.

Silverman, Daniel. 1992. "Multiple scansions in loanword phonology: evidence from Cantonese". *Phonology* 9. 289–328.

Smith, Jennifer. 2005. "Loan Phonology is not All Perception: Evidence from Japanese Loan Doublets". *Japanese/Korean Linguistics 14* ed. by Timothy J. Vance. Stanford: CLSI.

Stevens, Kenneth N. 1972. "The quantal nature of speech: Evidence from articulatory-acoustic data". *Human Communication: A Unified View* ed. by Eduard E. David & Peter B. Denes, 51–66. New York: McGraw-Hill.

Stevens, Kenneth N. 1989. "On the quantal nature of speech," *Journal of Phonetics* 17. 3–46.

Stevens, Kenneth N. 1998. *Acoustic Phonetics.* Cambridge, MA: MIT Press.

Stevens, Kenneth N. 2002. "Toward a model for lexical access based on acoustic landmarks and distinctive features". *Journal of the Acoustical Society of America* 111 4. 1872–1891.

Stevens, Kenneth N. & Blumstein, Sheila. 1978. Invariant cues for place of articulation in stop consonants, *Journal of the Acoustical Society of America,* 64, 1358–1368.

Stevens, Kenneth N. & Blumstein, Sheila. 1981. "The search for invariant correlates of phonetic features." *Perspectives on the study of speech* ed. by Peter D. Eimas & Joanne L. Miller. Hillsdale, NJ: Erlbaum.
Sussman, Harvey, Fruchter, D., Hilbert, J., & Sirosh, J. 1999. "Linear correlates in the speech signal: The orderly output constraint." *Behavioral and Brain Sciences* 21. 287–299.
Tervaniemi, Mari, J. Saarinen, Petri Paavilainen, N. Danilova, & Risto Näätänen. 1994. "Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity". *Biological Psychology* 38. 157–167.
Ungerleider, Leslie G. & Mortimer Mishkin. 1982. "Two cortical visual systems". *Analysis of Visual Behavior* ed. by David J. Ingle, Melvyn A. Goodale & Richard J.W. Mansfield, 549–586. Cambridge, MA.
Valdman, Albert. 1973. "Some Aspects of decreolization in Creole French". *Diachronic, Areal, and Typological Linguistics. Current Trends in Linguistics, vol. II* ed. by Thomas Sebeok. The Hague: Mouton.
Van Leuween, Cees. 1998. "Perception". *A Companion to Cognitive Sciences*. ed. by William Bechtel & George Graham, 265–282. Oxford: Blackwell Publishing.
Vance, Timothy J. 1987. *An Introduction to Japanese Phonology.* Albany, NY: State University of New York Press.
Waldstein, Robin S. 1989. "Effects of postlingual deafness on speech production: implications for the role of auditory feedback". *Journal of the Acoustical Society of America* 88. 2099–2144.
Warren, Richard. 1970. "Perceptual restoration of missing speech sounds". *Science* 167. 392–393.
Watkins, K. E., Strafella, A. P., & Paus, T. 2003. "Seeing and hearing speech excites the motor system involved in speech production". *Neuropsychologia* 41. 989–994.
Werker, Janet F., & Chris E. Lalonde. 1988. "Crosslanguage speech perception: Initial capabilities and developmental change". *Developmental Psychology* 24. 672–83
Werker, Janet F., & Richard C. Tees. 1984. "Crosslanguage speech perception: Evidence for perceptual reorganization". *Infant Behavior and Development* 7. 49–63.
Wetzels, Leo. This volume. "Nasal harmony and the representation of nasality in Maxacalí: evidence from Portuguese loans".
Wilson, Margaret. 2001. "The case for sensorimotor coding in working memory". *Psychonomic Bulletin and Review* 8.44–57.
Wilson, Stephen M., Ayse P. Saygin, Martin I. Sereno, & Marco Iacoboni. 2004. "Listening to speech activates motor areas involved in speech production". *Nature Neuroscience*. 7. 701–702.
Woldorff, Marty G., Steven A Hackley & Stephen A Hillyard. 1991. "The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones".*Psychophysiology* 28.1. 30–42.
Yates, Aubrey J. 1963. "Delayed auditory feedback". *Psychological Bulletin* 60. 213–251.
Yip, Moira. 1993. "Cantonese loanword phonology and Optimality Theory". *Journal of East Asian Linguistics* 2. 261–291.