

# A psycholinguistic perspective on the acquisition of phonology

*Franck Ramus, Sharon Peperkamp, Anne Christophe, Charlotte Jacquemot, Sid Kouider and Emmanuel Dupoux*

*This paper discusses the target articles by Fikkert, Vihman, and Goldrick and Larson, which address diverse aspects of the acquisition of phonology. These topics are examined using a wide range of tasks and experimental paradigms across different ages. Various levels of processing and representation are thus involved. The main point of the present paper is that such data can be coherently interpreted only within a particular information-processing model that specifies in sufficient detail the different levels of processing and representation. In this paper, we first present the basic architecture of a model of speech perception and production, justifying it with psycholinguistic and neuropsychological data. We then use this model to interpret data from the target articles relative to the acquisition of phonology.*

## **1. Introduction**

One recurrent problem in linguistic and psycholinguistic research is that data are often taken to reflect one or two particularly prominent levels of representation or processing, whereas many more are potentially involved. Different datasets sometimes appear to be contradictory because they are thought to reflect incompatible properties of one particular level of representation. But a full analysis of the tasks and of the levels of representation involved may reveal that such datasets actually reflect properties of *different* levels of representation. Therefore they are not necessarily in conflict, but the theoretical interpretation of the data may need to be revised.

In this paper, we discuss what we call the “standard model” of phonological theory. This model basically distinguishes two levels of mental representation, an underlying and a surface level. The former is a level of representation in which words are represented as a sequence of abstract units (phonemes, features . . . ),

the latter is a more detailed level of representation in which complete utterances are represented as a sequence of speech sounds or phones. The phonological grammar mediates between these two levels, in that it maps underlying forms onto surface forms.

In the times when spontaneous speech corpora were the main source of data to be interpreted, the standard phonological model was perhaps sufficient to account for most of the phenomena of interest<sup>1</sup>. But in the days of 21<sup>st</sup> century Laboratory Phonology, where relevant data is sought from perception as well as production tasks, first and second language acquisition, and where the analysis of phonetic details in speech production as well as of their influence on perception has become much more sophisticated, this model has clearly become insufficient. Basically, there is nothing wrong with the standard model, but it needs to be complemented with additional components in order to account for the extended range of available data, as some have recognised before (Hume and Johnson 2001; Boersma 1998, 2006; Pierrehumbert, Beckman, and Ladd 2000). Furthermore it is also crucial to take into account how the postulated information-processing architecture might plausibly develop from the initial stage at birth to the mature stage.

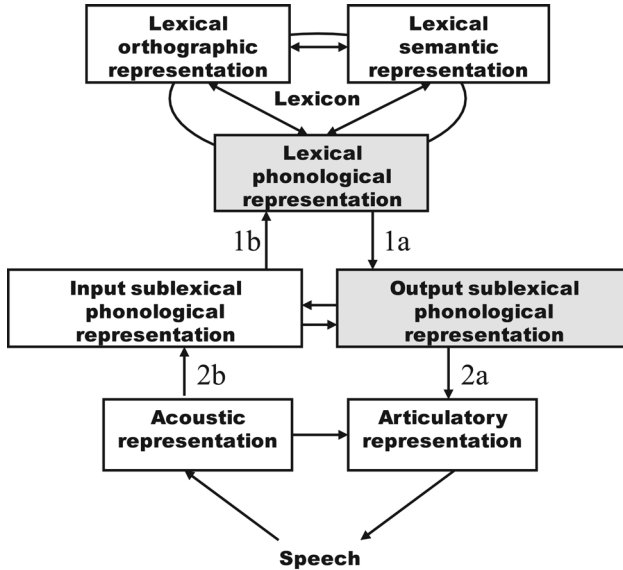
In this paper, we first present the basic architecture of a model of speech perception and production, justifying it with psycholinguistic and neuropsychological data. We then use this model to interpret data from the target articles relative to the acquisition of phonology.

## **2. An information-processing model of speech perception and production**

### **2.1. Basic architecture and functioning**

The model presented in Figure 1 is directly inspired from the classic logogen model (Morton 1969) and subsequent updates, variants and refinements (Norris, McQueen, and Cutler 2000; Morton 1980; Caramazza 1997; Coltheart 1978; Levelt 1989), as well as from ideas coming from the linguistic literature (Chomsky and Halle, 1968; Jackendoff, 1997; Prince and Smolensky, 1993).

The basic principles at work in Figure 1 are as follows: 1) boxes stand for distinct levels of representations, 2) arrows stand for “processes” that perform a mapping (or conversion, or translation) between different levels of representations, and 3) not all conceivable boxes and arrows are shown, only those that are necessary for the present discussion.



*Figure 1.* An information processing model of speech perception and production. The boxes in grey represent the standard model of phonological theory. Arrow 1a corresponds to output phonological processes, 2a to output phonetic implementation. 1b corresponds to phonological parsing (inverse phonology) and 2b to perceptual phonetic decoding. In the adult, these four processes are finely tuned to the phonological and phonetic properties to the maternal language(s). They may be mistuned during (first or second) language acquisition, or in cases of brain lesion or learning disability.

The model is centred around the mental lexicon, which is a long-term memory store divided into at least three parts that interface with different aspects of the world: semantic representations, phonological representations (including their segmental content and stress or tonal pattern) and orthographic representations. (Lexical morphosyntactic properties are not represented here, and syntax is more generally completely out of this picture).

Here only the adult state is represented. In the initial state, the overall architecture may be in place, but the lexicon is empty, representations are in a universal format untainted by any language-specific category, and similarly, processes are not trained to perform any language-specific function.

All the way from the cochlea to the primary auditory cortex, speech sounds, like all other sounds, are encoded in a non-specific manner: this is embodied by our *acoustic representation*. At a later stage of processing, speech must be

encoded in a speech-specific manner: this is a *sublexical phonological representation*. The arrow between the sublexical phonological representation and the phonological lexicon represents auditory word recognition.

Speech production includes the selection of the appropriate words (typically at the semantic level), the retrieval of their phonological form (from the phonological lexicon), their assembly into a whole phonological utterance (at the sublexical phonological level), and the conversion of this latter level into an articulatory representation that will trigger the motor commands producing speech (Levelt 1989).

It is evident that the standard phonological model consists of two of the boxes represented here (in grey in Figure 1), i.e. the lexical phonological representation and the output sublexical phonological representation, with arrow 1a going from the former to the latter representing phonological grammar. Embedding this standard model within the more comprehensive one highlights at least two other characteristics of the model: 1) there is an input pathway, distinct from the output pathway, but linked with it; 2) this input pathway is also subdivided between lexical, sub-lexical and peripheral (acoustic) levels. Before discussing the importance of these properties for the interpretation of laboratory phonology and language acquisition data, let us review some empirical evidence in favour of these various levels of representation.

## 2.2. Input versus output systems

There has been considerable debate in the speech processing community as to whether one should distinguish separate input and output speech systems or postulate a common amodal one. Of course, at the most peripheral level, auditory representations are separate from articulatory ones. The former are dedicated to the analysis of auditory patterns, and at that level, information consists of a continuous representation of sounds that include speech sounds as well as non-speech sounds. The latter is dedicated to the motor planning of articulatory gestures and consists in the specification of muscle movements/trajectories, which are adequate both for speech sounds and for other types of vocalizations.

The interesting question concerns the more abstract sublexical phonological level: should there be a single amodal phonological system or two separate ones? The strongest evidence for separate systems comes from neuropsychology, and in particular from cases of conduction aphasia. In this type of syndrome, patients have a relatively intact comprehension and speech production combined with a severe impairment in the ability to repeat speech (Caramazza et al. 1981). Jacquemot, Dupoux and Bachoud-Lévi (2007) explored the case of a patient (FA) who could perceive and produce both real words and nonwords, but who

could not repeat nonwords. Such a deficit can be accounted for in the model by positing that there are two distinct sublexical phonological representations, one for perception and one for production. Specifically, the incapacity to repeat non-words is evidence for an impaired link from the former to the latter (the repetition of real words is not affected, since the input and the output systems are also connected by the lexical phonological representation). A further assessment of FA also provided evidence for the existence of two separate links between the input and the output sublexical phonological representations, one in each direction. Indeed, FA had no problems with tasks that require to internally 'hear' phonological output (without overt production), such as judging whether two pictures' names rhyme, or whether a picture's name contains a previously heard target syllable. The results showed that FA had an intact conversion mechanism from phonological output to phonological input. Overall, these results strongly suggest that sublexical phonological representations in perception and in production are separate and connected by two independent links.

### 2.3. Levels of representations of speech sounds

In the output pathway, as classically analysed in the linguistic literature, the distinction between sublexical (surface) and lexical (underlying) representations stems from the detailed examination of the phonological shape of words. Due to a variety of phonological processes, a word can surface in a variety of ways depending on the phonological context, speaking rate, dialectal style, etc.. Such an architecture has been substantiated in the psycholinguistic literature: speech production models typically acknowledge this distinction (see Levelt 1989), and neuropsychological investigations have reported cases of specific impairments at either level. For instance, Goldrick and Rapp (2007) have recently reported two cases of production deficits where the patients had problems producing words, but had intact semantic and articulatory processes. In one patient, the errors were affected only by lexical factors such as lexical frequency and neighbourhood density, suggesting a deficit at the level of lexical representations. In the other patient, the errors were affected by phonological factors such as syllable position, place of articulation, and phoneme frequency, suggesting a deficit in sublexical phonological representations.

There is less consensus concerning the exact format of the two levels of representation. At the lexical level, whereas predictable variations are generally assumed to be derived by the grammar rather than being encoded underlyingly, the degree of abstractness of underlying representations is still a matter of debate (see Steriade 1995 for a review). At the sublexical level, the amount of phonetic detail argued to be present in surface representations varies. For some, there

is a separate grammar of phonetic implementation that maps phonological surface forms onto *phonetic* surface forms (Chomsky and Halle 1968; Prince and Smolensky 1993; Keating 1990, here arrow 2a). For others, by contrast, there is only one grammar and hence one level of surface representation (Flemming 2001). According to the model one adheres to, phonetic detail is thus either present at the surface phonological level or introduced at a later stage. In this paper we stay neutral with respect to this issue, and we simply assume that lexical representations are abstract in the sense that they do not include complete phonetic specifications of the word forms.

In the input pathway, the evidence for the distinction between acoustic and sublexical phonological representations is probably less well known within the linguistic literature. It rests on experiments demonstrating language-specific effects in the processing of speech sounds. In perception, hearers show considerable difficulties in discriminating and memorizing non-native contrasts. For instance, Japanese listeners have persistent trouble discriminating between English /r/ and /l/ even if these two phonemes are acoustically discriminable (Goto 1971; Lively et al. 1994). Such language-specific effects are not limited to segmental contrasts, but extend to suprasegmental regularities (see, for instance, Dupoux et al. 1997; Dupoux et al. 1999; Dupoux et al. 2001). The current interpretation of these effects is that experience with native categories shapes sublexical phonological representations (Best and Strange 1992; Best 1995; Kuhl 2000) and that these representations are automatically activated when processing speech. In theory, all speech stimuli could be differentiated at the (non language-specific) acoustic level of representation. The fact that listeners seem to find this difficult in many conditions suggests that, most of the time, they irrepressibly activate sublexical phonological representation and fail to attend to acoustic information that is not used contrastively in their native language. Language specific effects with speech sounds are therefore a strong reason to endorse the dissociation between acoustic and sublexical phonological representations.

This interpretation is again bolstered by neuropsychological data from aphasic patients with no hearing impairment, who can be selectively impaired in phonological processing (such as phoneme discrimination or identification tasks), while they are not impaired in tasks involving non-speech sounds (Metz-Lutz and Dahl 1984; Caramazza, Berndt, and Basili 1983; Auerbach et al. 1982). Moreover, in neuroimaging studies, it has been shown that phonological, but not acoustic processing, involves specifically the activation of the left planum temporale and supramarginal gyrus (Jacquemot et al. 2003; Dehaene-Lambertz et al. 2005).

#### 2.4. Consequences for theories of grammar and language acquisition

The usual notion of phonological grammar refers to the processes converting lexical phonological forms into output sublexical phonological forms. It is to be noted however that linguists and psycholinguists may differ on what counts as a phonological process. Linguists sometimes include in the phonological grammar, processes that are productive synchronically as well as processes that arose diachronically but are no longer active. For instance, phonological variations across morphologically related words (e.g., opaque/opacity, cf. Chomsky and Halle 1968) may sometimes only reflect statistical regularities present in the lexicon, rather than grammatical processes per se (Myers 1999). Such cases may reflect grammatical processes that occurred in the brains of yesterday's speakers of the language, and that have left their mark on the shape of today's lexicon. For psycholinguists however, synchronic and diachronic processes are very different: only the former require an active mental operation and need to be acquired by the child as such, whereas the latter just reflect the content of the lexicon.<sup>2</sup>

Whereas lexical regularities are sometimes unduly taken to reflect phonological grammar, another kind of phonological grammar is most often overlooked: the one that applies in the input pathway (arrow 1b). Indeed, phonological variations introduced by speakers are typically not noticed by listeners during online speech perception, although the same listeners are typically able to hear the differences when excised out of their context. This suggests that during on-line speech perception, there is a mechanism ("inverse phonology") that undoes at least some of these variations in order to facilitate the recognition of lexical forms. Empirical evidence for such a mechanism has been provided for assimilation processes, showing that English- and French-speaking listeners do mentally undo place and voice assimilations, respectively, when hearing words in assimilatory contexts (Darcy, Peperkamp, and Dupoux 2007; Gaskell and Marslen-Wilson 1996; Darcy et al. 2009). There are therefore two phonological grammars, one in the output pathway (figured by arrow 1a from lexical to sublexical representations), and one in the input pathway (figured by arrow 1b from sublexical to lexical input representations). The latter is unfortunately much less described than the former (but see Eisner 2002; Boersma 1998, 2006). Nevertheless it is an integral part of what the child has to learn.

In such a model, the input and output phonological grammars (1b and 1a) are theoretically distinct entities. We assume that, as a first approximation, these two grammars develop in a parallel fashion in children (through input-output loops) and end up being undistinguishable in monolingual adults. However, in cases of abnormal development, or in cases of second language acquisition, it

is possible that input and output phonology diverge. For instance, it has been observed that late bilinguals can sometimes show better production of a foreign contrast than its perception (Sheldon and Strange 1982). Another example is the Japanese vowel epenthesis process in the perception of foreign words, that has arguably no counterpart in the synchronic production grammar (Dupoux et al. 1999). One should note, in addition, that the arrows going from sublexical output to articulation (2a) and from acoustics to sublexical input (2b) represent processes that are not language independent; rather they involve categorization or planning processes which are finely tuned to the phonological categories of the native language (Kuhl et al. 1992).

To sum up, we argue that there are two phonological grammars, one in input and one in output, and two more peripheral (acoustic and articulatory), yet language-dependent, mapping processes. An alternative way to think about it would be that phonological grammar is distributed in a partly redundant way over several processing loci (perception, production, and at several levels). This may seem disconcerting from a linguistic perspective, but in fact redundant and overlapping processing systems are common in psychology and neuroscience.

Finally, the phrase “the acquisition of phonology“ may mislead one into thinking that there is just one thing to be acquired by the child, namely phonology. But from the previous discussion it becomes evident that there are different components to be acquired. One component is the right format of representation at each of the three levels of phonological representation (input and output sublexical, and lexical). As we will see below, it is not entirely clear whether the three levels are acquired more or less in parallel or whether lexical and output levels seriously lag behind the input sublexical level. A second component to learn is the lexicon itself. Two further components are both the input and the output sublexical phonological grammars. Here again, whether there are two distinct grammars to acquire, or whether one is the mirror image of the other is not clear. Acquiring phonology is therefore a multi-faceted problem for the child. Of course, all these components are not entirely independent from each other. But neither do they completely follow from one another. Let us now look more closely at the language acquisition data and assess what they imply for the acquisition of each component.

### **3. Interpreting language acquisition data**

The papers by Fikkert and by Vihman report data from both speech perception and production by infants and toddlers. Their tacit assumption is that the entire body of data can be taken to reflect the development of a single level of repre-

sentation, that is, the lexical phonological representation. But we will see that it may take more than one level of representation, and a careful consideration of the tasks used to generate the data, to provide a full account.

### 3.1. Speech discrimination

Starting with Eimas et al. (1971), many studies on speech perception by infants have converged on the idea that humans are born with certain universal auditory categories, possibly shared with other species, and that these categories change under exposure to the sounds of a particular language, reaching a relatively stable, language-specific state around the first year of life (Werker and Tees 2005; Kuhl 2000). In terms of our model, this represents a developmental tuning of the format of the input sublexical representation<sup>3</sup>. Indeed, in these speech perception tasks, infants typically discriminate between novel or pseudo-words, i.e., they compare two phonological forms represented at the input sublexical level. Furthermore, many experimental studies have shown that this tuning goes beyond phonemic categories, and that it also includes statistical information about the typical shape of words in the native language (stress and tonal pattern, phonotactics, etc. Jusczyk, Luce, and Charles-Luce 1994; Jusczyk, Houston, and Newsome 1999; Saffran, Aslin, and Newport 1996; Jusczyk, Cutler, and Redanz 1993; Mattock and Burnham 2006; Friederici, Friedrich, and Christophe 2007). Note that this phonological information about the lexicon is acquired during the first year of life before any significant number of lexical entries are acquired. Whereas this set of results suggest an early tuning of the input sublexical representation, it does not inform us on the development of lexical and output sublexical phonological representations.

### 3.2. Word learning

Infants acquire some of the cues helpful for word segmentation around age 9 months, and start to acquire their lexicon at the end of the first year of life (Jusczyk 1997). Considerable controversy has arisen regarding the format of these first lexical entries. While the default assumption would seem to be that the lexicon is based on the same format as that acquired during the first year of life in input sublexical representations, some researchers have posited either a much more detailed representation (including acoustic details, e.g. Singh, White, and Morgan 2008), while other researchers have posited a much less detailed (underspecified) representation (e.g., Fikkert, this volume). This remains a complex issue because of the experimental difficulty of unambiguously tapping the lexical level of representation without being contaminated by methodological confounds.

Starting with Stager and Werker (1997), many studies have tried to teach babies a novel word, and subsequently test how they react to a change in the phonological shape of the word. In this “switch paradigm”, infants are familiarized with the pairing of a novel word and the picture of an object. After habituation, infants are presented with the same object, either together with the same word (the same condition) or with a minimally different word (the switch condition). The difference in looking time between same and switch conditions is taken to indicate the degree of mismatch between the newly learnt lexical entry and the minimally different word. This task therefore provides evidence as to how much phonetic detail is encoded when storing a novel word for the first time in the lexicon, and how much can immediately be retrieved from this preliminary representation to be compared with a new item. A bottleneck at either stage is likely to limit performance. Furthermore it is quite a demanding task, in particular in terms of attentional resources, since it supposes a very fast encoding (7 to 10 presentations) of a novel item.

The main result obtained by Stager and Werker (1997) and subsequently replicated and extended (e.g., Pater, Stager, and Werker 2004), was somewhat counterintuitive: at the age of 14 months, infants failed to notice a minimal change in the word form, e.g. from ‘bin’ to ‘din’ (even though they were perfectly able to distinguish between ‘bin’ and ‘din’ in a speech discrimination task). However, they performed well when the novel words were very different, such as ‘lif’ and ‘neem’. By the age of 17 months, infants performed well in the word-learning task, even with minimal pairs of novel words. In their original paper, Stager and Werker proposed that infants who just begin to learn words may not pay attention to fine phonetic detail (even though they are able to perceive and represent them). They suggested that this inattention may be beneficial to the infant, in that it would free attentional resources for the task of word-learning itself. A more mundane variant of this interpretation, inspired by the task analysis above, is simply that the word learning task, as implemented by Stager and Werker, is much more complex and demanding than a discrimination task, and that various difficulty factors (speed, attention, subtlety of phonetic differences) cumulate, hence the decrease in infants’ performance, specifically in the minimal change condition of the word learning task.

Another interpretation of these results would be that infants’ early lexical representations are not fully phonetically detailed. This is the interpretation espoused by Fikkert (this volume). On her account, the place of articulation feature is unspecified for ‘coronal’. Furthermore, babies would initially (at 14 months) represent only one place of articulation feature per word, that of the stressed vowel. This proposal makes very specific and somewhat counter-intuitive predictions as to the infants’ behaviour in the task. For instance, Fikkert predicts that

in the switch paradigm, 14-month-old infants should not distinguish between 'din' and 'don'. She also predicts that infants should distinguish 'bin' from 'bon' only when 'bon' is presented first (and thus stored lexically). Nevertheless both predictions are borne out by the results of Fikkert's experiments.

But underspecification theory makes even more problematic predictions. For instance, since 'bin' is encoded as 'Ø' as far as place of articulation is concerned, it does not mismatch with anything else; thus, a child habituated with 'bin' should not notice a change even to a very different word, say, 'gom'. Some earlier studies have contrasted words that are very different, i.e. 'lif' and 'neem'. In those experiments, infants do notice the difference (Werker et al. 1998; Stager and Werker 1997). However, these words differ by other features than place of articulation, so they are not appropriate for this purpose. Therefore, a more specific test of this prediction would be warranted.

Another odd consequence of the underspecification hypothesis is that 'don' (stored lexically as labial) mismatches with itself (because when represented sublexically in the test phase, [d] is not labial). This predicts that 'bon' is recognized as a good instance of 'don', whereas 'don' itself is not, as described in Fikkert's Table 8. Fikkert concludes that in the word learning task, infants should look longer to the "switch" item (the one different from the habituation item). However, it seems to us that this is not the correct prediction. Table 8 indeed predicts that infants, when habituated to 'don', should look longer to the same item 'don' than to the switch item 'bon'; it is only when habituated to 'bon' that they should look longer to the switch item 'don'. Across the two habituation conditions, the two opposite effects should average out to zero. But this is not what is found.

If the interpretation of Stager and Werker's results in terms of 'attentional load' or 'task difficulty' is correct, then any simplification in the learning task itself should improve the infant's performance. Ballem and Plunkett (2005) showed that, using a preferential-looking (rather than habituation) procedure, 14-month-olds are able to differentiate newly learnt words differing by one phonetic feature (like 'tuk'–'puk'). Preferential looking is thought to be easier because babies have the two alternatives in front of their eyes, rather than having to appeal to memory. Thus, there is clearly an effect of task difficulty. When it is reduced, it appears that 14-month-old infants' may after all be able to make fine phonetic distinctions between novel words. This predicts that if Fikkert reran her experiments by using preferential looking rather than habituation, she might obtain positive results for all contrasts, confirming the full specification of 14-month-olds' early lexical representations.

To summarise, it seems to us that the best explanation of Stager and Werker's results on word learning is to be couched in terms of the difficulty factors that

bear on a given task at a given age, and therefore that it is not necessary to postulate lexical underspecification. Nevertheless, Fikkert's specific pattern of results obtained at 14 months with the habituation procedure cannot be explained in terms of difficulty factors. Are there alternative ways to explain this pattern of results? We can only point to potential methodological confounds.

For instance, take the asymmetry between 'bin'–'bon' and 'bon'–'bin' discrimination. This asymmetry is manifested by the fact that in the 'bon'–'bin' condition, infants look massively longer to switch item 'bin', while in the 'bin'–'bon' condition they fail to look longer to switch item 'bon' than to 'bin' (Fikkert, this volume). This could result from the addition of two independent effects: perfect discrimination between 'bin' and 'bon', and overall preference for 'bin'. Why would infants prefer 'bin'? We have no theory about that, but such preferences are commonplace and maybe related, among other things, to the statistics of the input. For instance, 9-month-old infants prefer to listen to more frequent phonemes or phonotactic patterns (Friederici and Wessels 1993). Regardless of what may drive infants' preferences, the bottom line is that the switch paradigm is not suited to testing asymmetries in discrimination, because it does not factor out intrinsic preferences. Baseline looking times for both words would be necessary to interpret the looking times in the switch and same conditions.

Regarding the finding that 'bin'–'din' is harder to discriminate than 'bon'–'don', this conclusion would first require a statistically significant interaction between pair and condition before being accepted. Secondly, before interpreting such a pattern of data, one would like to be sure that both pairs are equally discriminable from an acoustic or phonetic point of view. This might be true in general for these pairs of syllables, or might be an artefact of the material used in these experiments. Indeed, any stimulus set that relies on a limited number of tokens of each type, uttered by a single speaker, runs the risk that the items might be discriminated on the basis of some phonetic details that might be completely idiosyncratic to that particular speaker or even to those particular recordings. A more varied stimulus set, using numerous tokens uttered by several speakers, would rule out this possibility, forcing discrimination at an abstract phonological level of representation, therefore ensuring that task performance actually reflects the intended representations. Regardless of their source, the hypothesis that asymmetries might arise from the material could be assessed by conducting acoustic measurements, or by testing the discriminability of these two pairs under various levels of noise.

These remarks regarding the data collected by Fikkert do not show that her conclusions are wrong, but that without additional data, simpler interpretations are possible. Importantly, these interpretations are motivated by the architecture in Figure 1. Preference for certain phonological shapes arise from statistics

collected by infants at the level of the input sublexical representation. Acoustic or phonetic effects on the discriminability of pairs of words or nonwords rest on the fact that many experimental tasks can be performed at several processing levels. A so-called 'lexical' task can involve sublexical or acoustic components; an 'input' task can involve output components, etc. This is because the processing system is designed to activate automatically all the levels, not just the ones that are of interest to the experimenter.

### 3.3. Word recognition

Beyond word learning tasks, a more direct way to address the phonetic specification of lexical representations is to test the representations of those words that are already familiar to babies. In such experiments, babies are typically presented with two pictures of familiar objects (say a car and a ball), while hearing sentences like 'look at the X!' where X either matches a lexical phonological form ('ball'), or introduces a minimal mispronunciation ('gall') (Swingley and Aslin 2000). Results show that English-learning babies, as young as 14 month-old, look faster to the target when it is correctly pronounced, than when it is minimally mispronounced (Swingley and Aslin 2002), suggesting again that their lexical representations do encode this degree of phonetic detail (see also Fennell and Werker 2003). Note that these results are completely at odds with Fikkert's assumption that 14-month-olds do not represent the place of articulation of consonants.

However, as Fikkert notes, the underspecification hypothesis makes additional predictions for word recognition that are not directly tested in published work (since the above authors did not compare the appropriate contrasts). Fikkert (this volume) also ran word recognition experiments with 20- and 24-month-old Dutch infants, but introduced one additional manipulation: she contrasted cases in which the target word started with a coronal consonant, such as in 'tand' (*tooth*), supposedly underspecified for place of articulation, and cases in which the target word started with a labial consonant, such as in 'poes' (*cat*), supposedly specified for place of articulation. She showed that infants accepted a change in the place of articulation of the first stop consonant, only when that consonant was coronal, but not when it was labial. These results are clearly consistent with the underspecification hypothesis, and can hardly be explained by a different theory. This is particularly surprising since one would expect that the lexical representations of 24 month olds would be even better specified than those of 14- and 17-month olds.

Again, if one pushes the logic of the underspecification theory to its limits, one inevitably makes strange predictions. For instance, let us take a word such

as ‘did’<sup>4</sup>. According to Fikkert, its place of articulation is encoded as ‘Ø’, even for older infants who can represent the place of articulation of all segments. As a result, many other potential words or pseudowords (like ‘big’, ‘dig’, ‘bid’, ‘gig’) should fail to mismatch with ‘did’. In other words, if 24-month-olds were put in an experiment in which one of the words was a fully underspecified one (such as ‘did’), they should always look longer to the picture for that underspecified word upon hearing “look at the X”, no matter what place of articulation features “X” carries. Wouldn’t that be very surprising? Wouldn’t that prevent them from learning new words? For instance, after having acquired the word ‘doll’, English-learning infants would be unable to learn the word ‘ball’, since it would never mismatch with the stored representation of ‘doll’: each new instance of ‘ball’ would be assimilated to ‘doll’. Yet there is good evidence that 14-month-olds have correctly specified representations for both ‘ball’ and ‘doll’ (i.e. they look longer to the correct picture when seeing the two pictures and hearing one of the words) (Fennell and Werker 2003). Is it because all the infants tested learned ‘ball’ before ‘doll’?

Obviously the pattern of results found by Fikkert is very intriguing and does not seem consistent with the hypothesis that infants’ lexical representations are fully-specified. Yet this same hypothesis is well supported by independent data, and the underspecification hypothesis makes a number of predictions that seem hardly tenable. Overall the entire set of data reviewed here seems inconsistent, which calls for a very close scrutiny of these data. Attention to all the methodological details, as well as a good information-processing model, are needed to try to understand exactly how, in each experiment, task structure might affect performance, stimuli might introduce biases, etc. Ultimately, only replications varying experimental procedures and stimuli will enable us to have a clearer picture of the factors driving infants’ performance.

### 3.4. Word production

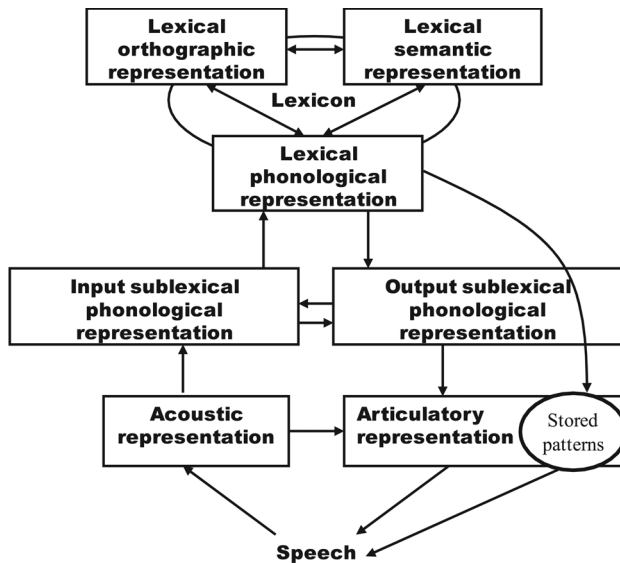
It has long been known that children’s performance in production lags behind that in perception. In particular, children perceive contrasts in adult speech that they neutralize in their own speech (see, for instance, Smith 1973). In order to account for the multiple errors and hesitations in young children’s productions, it has been proposed that children apply a set of phonological rules that are not part of the adult grammar. These rules take surface adult forms as their input and result in surface child forms, consisting of simpler phonological structures (Smith 1973). In the framework of child phonology, the acquisition of the adult phonological grammar thus largely consists of the gradual abolishment of these simplifying rules (or, in Optimality Theory, of the demotion of the rele-

vant markedness constraints). This phonological acquisition would take several years, beginning at around 12 months of age and lasting until around five or six years. This is the grammatical account of child phonology. Fikkert (this volume) makes a different proposal, which might be termed the representational account, according to which children's productions are limited mainly by the features that they can represent in their phonological lexicon. Alternatively, a quite different hypothesis to account for children's production data is that children rapidly converge towards the adult phonological lexicon and grammar, and that deviations from the adult targets merely reflect the development of the articulatory representation and the link thereto (1b) from the output sublexical phonological representation (Faber and Best 1994). Vihman (this volume) advocates this articulatory account, while at the same time holding to the idea that early lexical representations are different from adults' (by being "holistic").

The articulatory development hypothesis may account for several characteristics of young children's productions that are otherwise hard to explain. For instance, acquisition is relatively slow and the changes in children's productions are gradual. Indeed, articulation is a very complex motor skill that requires the fine coordination of some 150 muscles in order to program and realize more than ten phonetic targets per second. Like any other motor skill, articulation is learned progressively. Another feature of children's productions is that frequent words are pronounced incorrectly for a longer period than infrequent words (Gierut and Storkel 2002). For instance, children often preserve immature forms for some words (say, French "*trou*" 'hole' as [kʁu]) long after they have overcome the articulatory difficulty, as shown by their correct production of other, less frequent, words ("*tronc*" 'trunk' as [tʁɔ̃]). This inverse frequency effect is difficult to explain both on the representational and on the grammatical accounts, since on the contrary, the more frequent a word is, the easier it should be to acquire its correct representation, and the more evidence it provides for the modification of phonological grammar.

Under the articulatory account, however, the effect can be explained by a feature of the speech production system that was proposed independently for adults: there are in fact two routes for establishing articulatory plans: a regular, assembly route (that shown in Figure 1), and a route that retrieves stored plans for frequent patterns (added in Figure 2). This was the "phonetic syllabary" or "gestural score" in Levelt (1992). Here we propose that the gestural score may include not only syllables, but also whole word forms, at least those most frequently pronounced. This may indeed help explain some word-specific idiosyncrasies in adult speech. Concerning child speech, the idea is that the early words that are pronounced frequently are stored in the gestural score. Because the child's articulatory skills are immature, the words are stored in the

simplified form that these skills initially allow. While articulatory skills remain limited, the child will continue to utter this simplified form, thereby reinforcing the stored plan. The child can hear his/her mispronunciation, thereby getting negative feedback that will ultimately drive the modification of the stored plan. But if the word was frequently uttered and therefore strongly reinforced, it may take a lot of negative feedback, even after articulatory skills have improved, to correct the stored pattern.



*Figure 2.* A modified information-processing model including the direct articulation route to the gestural score containing the stored patterns of the most frequent word forms.

Our account of children's early production implies that it is very difficult to interpret the source of a given deviation from the adult target: Is it a stored pattern? Or does it reflect the standard phonological route? And, in the latter case, does it reflect an immature phonological representation, or grammar, or articulation? The multiplicity of factors affecting children's surface forms, and the traditional focus on grammar as the only factor of interest, means that we actually know very little about the development of phonological grammar and of output sublexical phonological representations. Child phonologists will have to be methodologically creative if they are to tease apart the different factors.

Let us consider, for instance, children who neutralize a certain contrast in production (e.g., Smith 1973). We have already argued that they must neverthe-

less have adult-like input sublexical and lexical phonological representations. How about their output sublexical representations? According to the articulatory account, they might well be adult-like too. A way of testing this hypothesis would be to ask them to make judgments on the form of words. For instance, children who merge [θ] and [s] could be shown a picture of a mouth and one of a mouse and asked which one rhymes with the word “house”. The words “mouth” and “mouse” not having been produced by the experimenter, the children will have to produce them internally (in their output sublexical representation), and compare them with “house”. This task therefore taps the output sublexical phonological representations while bypassing articulation. The problem with this kind of tasks is that it requires metaphonological awareness, a capacity that develops relatively late in children (4 to 6 years, depending on the unit to be judged, Duncan et al. 2006), perhaps too late to test any interesting aspect of child phonology. Again, indirect methods such as priming might prove fruitful (in the present case, one would be tempted to try picture-pseudoword interference, Schriefers, Meyer, and Levelt 1990). However, this method requires averaging over many trials per condition, so it will never provide information about the representation of any particular item, but only of a set of items.

Consistent with the articulatory account, but also taking a broader view, Vihman (this issue) asks more generally which factors explain variance in children’s word forms, including variance between children’s forms and the adult targets, variance among languages, and variance among children learning the same language. Let us expand slightly on her answers. Children’s word forms are shaped by:

- 1. Universal factors.** In particular, universal factors constraining the development of motor control of the articulators are the main source of differences between children’s and adults’ words. Quite possibly, additional sources are universal linguistic factors (markedness), to the extent that they can be shown to be irreducible to motor constraints. The use of a unique place of articulation over the whole word, as described by Fikkert (this volume) in the earliest stage, is a good illustration of a putative linguistic constraint that can plausibly result from a universal articulatory constraint (it is difficult to rapidly change the place of articulation).
- 2. Language-specific factors.** As reviewed earlier, the ambient language rapidly shapes the child’s input sublexical and lexical representations. This must be the main source of resemblance between the word productions of children acquiring the same language: the targets are the same, and are correctly represented. For the same reason, these representations are not likely to explain much variance between children’s and adults’ words. The development of the

output phonological representations and grammar, and therefore the contributions of these components, are largely unknown, given that they have not been investigated independently of articulatory constraints.

- 3. Idiosyncratic factors:** what specific words the child has heard, what s/he wants to say... It is also plausible that, on top of universal articulatory constraints, children may vary slightly in the respective control that they have over their tongue, their lips, their larynx, and in sequential planning capacities, etc., so that these child-specific articulatory constraints may drive the predominant use of particular word forms. Such factors would therefore be the main source of variance between children learning the same language.

Vihman's (this issue) paper is largely focused on the last point. Her data lead her to postulate a two-stage mechanism, including first the acquisition of a few individual items ("holistically" represented), and subsequently the selection of a subset of these forms to abstract out a template, that will thence be used to adapt all word forms. We should caution that nothing in the data specifically indicates the existence of two stages, or the systematic use of a real template. The simple assumption that children speak under both universal and individual articulatory constraints is sufficient to account for all the data presented by Vihman. Therefore, although we largely agree with her about the main sources of variance in children's speech, we take the two-stage mechanism and the template hypothesis to be metaphorical rather than explanatory.

Finally, the assumption that lexical representations are initially holistic is neither warranted, nor necessary to account for production data (in fact it plays no role in Vihman's argument). As we have seen before, it is inconsistent with perceptual and word learning data. And since children's productions are mainly constrained by limits on articulation, they say little about the format of phonological representations. The very notion of a holistic representation may in fact be fundamentally flawed: what exactly could it mean for a representation to be holistic? What could be produced from a holistic lexical representation, if not something totally slurred with no identifiable parts<sup>5</sup>? Children's productions may well be off the adult targets but they are anything but slurred.

#### **4. Episodic memory**

Episodic memory is of interest to laboratory phonologists insofar as it may affect performance in some of the tasks they use. Such an excursion may in particular be useful to interpret the data obtained by Goldrick and Larson (this volume). Episodic memory refers to the memory of specific events and of all

the representations associated with these events. Whenever we hear a word, we not only access the lexicon as indicated by the model in Figure 1, but we also process the identity of the speaker who uttered it, his/her voice, his/her prosody, the emotional state this prosody conveyed, the context in which the word was uttered (time, place, situation), etc. All of this is encoded into the memory trace for this particular episode.

In psycholinguistic research, interest in episodic memory rose steeply when it appeared that episodic memories associated to a word could subsequently affect its retrieval. In a typical recognition memory experiment, subjects first undergo a study session where they hear (and for instance type to dictation) a long list of words uttered by several speakers, then in a subsequent test session (that can take place as much as a week later), they hear a list of both old and new words, and must decide for each word if they had heard it in the first session or not. It appears that their recognition performance in the test session is increased for words uttered by the same voice as in the study session, than for words uttered by a different voice (Goldinger 1996). Furthermore, there is good evidence that it is not just the repetition of speaker identity that does the priming, but the repetition of acoustic properties of the words, as there are gradient effects of acoustic similarity (Goldinger 1996). Such results have led some authors to propose that the phonological lexicon, rather than being made of abstract phonological representations, is made of episodic memories, containing in particular all the acoustic details of words as they are heard (Goldinger 1998; Johnson 1997; Pierrehumbert 2001).

However, it turns out that voice effects are task-dependent. They appear in tasks involving explicit memory recall of these traces, but not in more implicit tasks. For instance, in a similar paradigm as before, including a study and a test session, voice effects disappear when the task in the test session is lexical decision<sup>6</sup> rather than lexical recognition (Luce and Lyons 1998). This is a problem for episodic lexicon models, because in these models acoustic details cannot be bypassed: they are the stuff that lexical entries are made of. In another study, Kouider and Dupoux (2005) investigated auditory subliminal priming, in which subjects perform a lexical decision task on a target word, preceded by a prime word that is masked so as not to be consciously heard. Even under those totally implicit conditions, a repetition priming effect is obtained, that is, reaction times for the lexical decision are shorter when the prime is the same word as the target, than when it is entirely unrelated. Here, an episodic model of the lexicon would predict a greater priming effect when prime and target are uttered by the same voice than by different voices. Yet just the same amount of priming was found, as predicted by an abstract model of the lexicon (Kouider and Dupoux 2005).

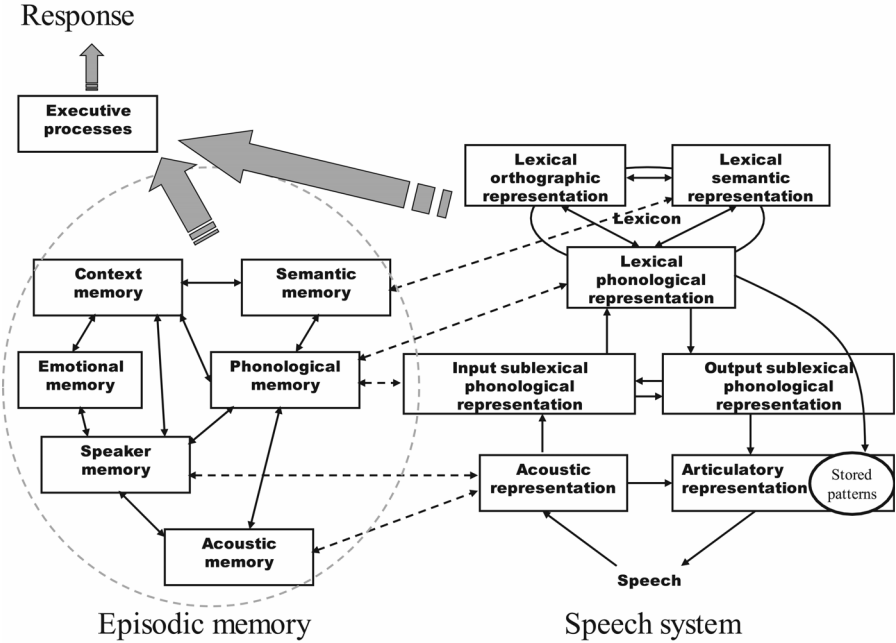


Figure 3. A modified information-processing model of the speech system including episodic memory and executive processes. This graphical representation is for illustration purposes and does not intend to make any claim about the format and structure of episodic memory.

A plausible explanation for these contrasting results is that explicit memory tasks like lexical recognition focus the subject’s attention on episodic memory (“did I hear that word in the first session?”), while implicit memory tasks like lexical decision don’t, and instead incite the subject to respond purely on the basis of lexical status. Given that there is evidence for both abstract lexical effects, and effects of episodic memories, it seems that the model that may best account for the whole data set is simply a model that includes both an abstract lexicon, and an episodic memory (Figure 3). Obviously the notion of an abstract lexicon never was incompatible with episodic memory, which everybody knows must exist for independent reasons. Interestingly, some proponents of the episodic lexicon have recently made significant steps in this direction (Pisoni and Levi 2007; Goldinger 2007).

Another important component that the model in Figure 3 now shows explicitly concerns the executive processes. By this we mean the cognitive function whose role it is to receive input from various modules, from episodic and from

long-term memory, to control the execution of the task (in the context of a psychological experiment), and to make decisions as to which behaviour to output<sup>7</sup>. Executive processes are implicit in all cognitive models, but sometimes their absence from the graphical representation leads one to forget that the behaviour in any task is not directly driven by the internal representations that are the target of the task. Now the model makes it obvious that responses in a given task can be potentially influenced by many different representations, be they in the speech system, in episodic memory or elsewhere. Responses can also be influenced by task-specific strategies. More generally, task structure biases which representations have the greatest influence on executive processes, hence on behavioural responses. Thus, a lexical recognition task, which incites subjects to explicitly search their episodic memory, allows the content of acoustic episodic memories to influence responses. A lexical decision task, which is better performed by searching the lexicon, and which would in fact be hindered by paying attention to episodic memories, leaves the latter little influence on responses. Hence the task-dependence of voice effects.

Beyond the debate between the episodic and the abstract lexicon, the broader view afforded by the model in Figure 3 also provides an alternative way to interpret tasks such as the one used by Goldrick and Larson (this volume). These authors used a task drawn from a common class of paradigms, in which subjects are exposed repeatedly to stimuli presenting a certain statistical pattern (in the present case a biased distribution of some segments across syllabic positions), and consequently manifest an implicit change in their behaviour in the direction of that statistical pattern (here, they produce speech errors that tend to match the biased distribution of the stimuli).

Goldrick and Larson use this paradigm to ask questions about what patterns are learnable (or not) in phonological acquisition. Some caution is in order, however, when using a ten-minute exposure in adults (who have already acquired a phonological system) to model an acquisition process that extends over several years in children. It is worth considering the possibility that part of the observed results in the adult case could come from the episodic memory system, rather than from the phonological representations per se.

This is not to deny that phonological learning is possible in adult speakers, or to argue that the Goldrick and Larson's results (this volume) are necessarily episodic. But these results have to be compared to others that show that the phonological system tends to resist late influences, even lasting for decades. For instance, even after extensive training in changing the perception and production of some non-native contrasts (like /r/ vs /l/ in Japanese learners of English) performance remains non-native (Lively, Logan, and Pisoni 1993). Exactly under what conditions and to what extent the phonological system may change

remains to be established (e.g., Darcy, Peperkamp, and Dupoux 2007; Dupoux et al. 2008; Sancier and Fowler 1997). In brief, our point is that when exposure is very brief, and test follows immediately (in the case of Goldrick and Larson training and test were simultaneous), it may be useful to consider the potential influence of episodic memory. Experimental manipulations like introducing changes in speakers, or increasing the lag between exposure and test can successfully reduce such influences (Fowler, Napps, and Feldman 1985; Kouider and Dupoux 2009).

To conclude, Goldrick and Larson are right that the simplicity of the phrase “probability matching” is deceptive. Indeed we have argued that it is not always clear what kind of learning probability matching reflects. However, the main point of Goldrick and Larson is that their subjects seem to be influenced by only some distribution biases. Indeed it is always interesting to observe that some patterns are easier to learn than others, as this may reflect cognitive constraints on the learning mechanisms. What remains to establish here, is whether the limitations observed in the learning of statistical patterns by adult subjects actually reflect constraints that bear on phonological acquisition (and a fortiori on phonological acquisition by the child). An alternative possibility would be, for instance, that certain syllables (say, those with /s/ in onset) are more easily articulated than others for purely motoric reasons<sup>8</sup>. Such an alternative hypothesis could be tested by assessing the baseline ease of articulation of the relevant syllables in the absence of exposure to a biased distribution of segments.

## 5. Conclusion

In this paper, we have taken the papers by Fikkert, Vihman, and Goldrick and Larson as case studies to analyse the many pitfalls that lie in the analysis of linguistic and psycholinguistic data, and to argue for the necessity to always analyse tasks within an information processing model in order to get clues about the processes and levels of representations that the data may reflect.

Chomsky (1976) claimed already long ago that linguistics was a branch of psychology. He was probably more right about it than most linguists (and perhaps he himself) have realised. Here we argue that the *methods* of linguistics should also be a branch of the methods of psychology. Indeed, all linguistic data are behavioural data. Linguistic representations are hidden in the brain, and can never be accessed directly by the experimenter. The experimenter can only observe the behavioural data, which bear some (complex) relationship with linguistic representations and grammar. Behavioural data are always collected using a task. All tasks involve multiple levels of representation and processing.

Data interpretation therefore relies on figuring out which levels the data reflect, hence on systematically distinguishing all the relevant levels of representation and the associated grammars. Not surprisingly, it often takes a great deal of methodological sophistication to design tasks that produce patterns of results that can be unambiguously attributed to a given representation or process. To hammer this point in, let us consider that even the simplest of all sources of linguistic data, speaking, is a task in a non-trivial sense. We have seen for instance in the case of child language how difficult it may be to attribute the observed patterns to a particular level of representation or processing. This should increase our awareness of the problems raised by the interpretation of data, and drive us to enrich the methodological repertoire by drawing from a greater range of tasks and by paying careful attention to the methodological conditions that may affect task performance.

### **Acknowledgments**

We wish to acknowledge feedback from Kie Zuraw, and financial support from the European Commission (Neurocom project) and the Agence Nationale de la Recherche (ANR-05-BLAN-0065-01). Correspondence to: Franck Ramus, LSCP, 29 rue d'Ulm, 75005 Paris, France. Email address: [franck.ramus@ens.fr](mailto:franck.ramus@ens.fr).

### *Notes*

1. This is only true up to a point: speech corpora must be listened to and coded, which inevitably recruits the listener's own speech perception pathway, phonology, and metalinguistic abilities.
2. Productivity remains a key criterion of grammatical processes. However, explicit generation or judgement of morphological forms may not be a sufficiently stringent test of productivity. It is likely that phonological, like syntactic grammatical processes, are largely inaccessible to consciousness. When explicitly asked to generate or judge a derived form, subjects may simply respond what *they think* is correct, based on a rapid survey of similar examples in their mental lexicon. Therefore it may not be surprising if such tasks yield results globally consistent with the statistics of the lexicon, but it is not clear that this reflects real grammatical processes. One may then wonder what kind of empirical data would constitute a sufficient proof of productivity. The problem of experimentally tapping subjects' unconscious mental processes without being contaminated by their conscious beliefs and strategies is a general one in psychology, whose solution is often to use more indirect methods where the experimental factors being manipulated cannot be detected by subjects. A well-known example is priming, where the relationship between a prime and a target modulates a behavioural response (usually reaction time), unbeknownst to the subject. In the present case, one might predict that, in a suitable experimental paradigm,

- word forms productively derived from one another would prime each other, whereas more superficially related word forms would not (Kouider and Dupoux 2009).
3. It may be improper to talk about a phonological representation in the initial state, but this may simply be understood as a higher-order auditory representation that adapts to the sounds of a particular language, hence becomes phonological.
  4. 'Did' may not be a very good example of a word very familiar to English-learning babies, and moreover it is not imageable so as to be used in a preferential looking experiment. But let us imagine that it were.
  5. As soon as just two parts could be somewhat reliably identified, the representation would not be holistic anymore. It might be underspecified, but not holistic (as Fikkert points out).
  6. Lexical decision involves deciding whether each item is a word or not, with typically half the items being pseudowords.
  7. To be entirely consistent, the model should show that the articulatory representation also feeds into executive processes, which control speech just as much as other behavioural output.
  8. In addition, we find that the evidence that subjects' production of /s/ in onset is less affected by the exposure than that of /s/ in coda or /f/ in either onset or coda is quite weak. A significant proportion  $\times$  condition interaction, as well as a replication would be in order before drawing conclusions on the special status of /s/ in onset.

## References

- Auerbach, Sanford H., Terry Allard, Margaret Naeser, Michael P. Alexander, and Martin L. Albert  
 1982 Pure word deafness. Analysis of a case with bilateral lesions and a defect at the prephonemic level. *Brain* 105 (Pt 2):271–300.
- Ballem, Kate D., and Kim Plunkett  
 2005 Phonological specificity in children at 1;2. *Journal of Child Language* 32 (1):159–73.
- Best, Catherine T.  
 1995 A direct realist perspective on cross-language speech perception. In *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*, edited by W. Strange, 167–200. Timonium MD: York Press.
- Best, Catherine T., and Winifred Strange  
 1992 Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20:305–331.
- Boersma, Paul  
 1998 *Functional phonology*. University of Amsterdam: Ph.D. dissertation.

- Boersma, Paul  
2006 A programme for bidirectional phonology and phonetics and their acquisition and evolution. In *LOT Summerschool*. Amsterdam.
- Caramazza, Alfonso  
1997 How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14 (1):177–208.
- Caramazza, Alfonso, Annamaria G. Basili, Jerry J. Koller, and Rita S. Berndt  
1981 An investigation of repetition and language processing in a case of conduction aphasia. *Brain and Language* 14 (2):235–71.
- Caramazza, Alfonso, Rita S. Berndt, and Annamaria G. Basili  
1983 The selective impairment of phonological processing: a case study. *Brain and Language* 18 (1):128–74.
- Chomsky, Noam  
1976 *Reflections on language*. London: Temple Smith.
- Chomsky, Noam, and Morris Halle  
1968 *The sound pattern of English*. New York: Harper and Row.
- Coltheart, Max  
1978 Lexical access in simple reading tasks. In *Strategies of Information Processing*, edited by G. Underwood, 151–216. London: Academic Press.
- Darcy, Isabelle, Sharon Peperkamp, and Emmanuel Dupoux  
2007 Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners. In *Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde, 411–442. Berlin: Mouton de Gruyter.
- Darcy, Isabelle, Franck Ramus, Anne Christophe, Katherine Kinzler, and Emmanuel Dupoux  
2009 Phonological knowledge in compensation for native and non-native assimilation. In *Variation and Gradience in Phonetics and Phonology*, edited by F. Kügler, C. Féry and R. van de Vijver. Berlin: Mouton De Gruyter.
- Dehaene-Lambertz, Ghislaine, Christophe Pallier, Willy Serniclaes, Liliane Sprenger-Charolles, Antoinette Jobert, and Stanislas Dehaene  
2005 Neural correlates of switching from auditory to speech perception. *Neuroimage* 24:21–33.
- Duncan, Lynne G., Pascale Colé, Philip H. K. Seymour, and Annie Magnan  
2006 Differing sequences of metaphonological development in French and English. *Journal of Child Language* 33 (2):369–399.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler  
1999 Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25 (6):1568–1578.

- Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi, and Jacques Mehler  
2001 New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 16 (5/6):491–505.
- Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian, and Jacques Mehler  
1997 A destressing “deafness” in French? *Journal of Memory and Language* 36:406–421.
- Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarrete, and Sharon Peperkamp.  
2008. Persistent ‘stress deafness’: The case of French learners of Spanish. *Cognition* 106:682–706.
- Eimas, Peter D., Einar R. Siqueland, Peter W. Jusczyk, and James Vigorito  
1971 Speech perception in infants. *Science* 171:303–306.
- Eisner, Jason  
2002 Comprehension and compilation in Optimality Theory. Paper read at 40th annual meeting of the Association for Computational Linguistics, at Philadelphia.
- Faber, Alice, and Catherine T. Best  
1994 The perceptual infrastructure of early phonological development. In *The Reality of Linguistic Rules*, edited by R. Corrigan, S. D. Lima and G. Iverson, 261–280. Amsterdam: John Benjamins.
- Fennell, Chris T., and Janet F. Werker  
2003 Early word learners’ ability to access phonetic detail in well-known words. *Language and Speech* 46:245–264.
- Flemming, Edward  
2001 Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7–44.
- Fowler, Carol A., Shirley E. Napps, and Laurie Feldman  
1985 Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition* 13 (3):241–255.
- Friederici, Angela D., Manuela Friedrich, and Anne Christophe  
2007 Brain responses in 4-month-old infants are already language specific. *Current Biology* 17 (14):1208–1211.
- Friederici, Angela D., and Jeanine M. I. Wessels  
1993 Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics* 54 (3):287–295.
- Gaskell, M. Gareth, and William D. Marslen-Wilson  
1996 Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception & Performance* 22 (1):144–158.
- Gierut, Judith A., and Holly L. Storkel  
2002 Markedness and the grammar in lexical diffusion of fricatives. *Clinical Linguistics & Phonetics* 16:115–134.

- 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 (5):1166–1193.
- Goldinger, Stephen D.  
1998 Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105 (2):251–279.
- Goldinger, Stephen D.  
2007 A complementary-systems approach to abstract and episodic speech perception. Paper read at 16th International Congress of Phonetic Sciences, at Saarbrücken.
- Goldrick, Matthew, and Brenda Rapp  
2007 Lexical and post-lexical phonological representations in spoken production. *Cognition* 102 (2):219–260.
- Goto, H.  
1971 Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia* 9:317–323.
- Hume, Elizabeth, and Keith Johnson  
2001 A model of the interplay of speech perception and phonology. In *The role of speech perception in phonology*, edited by E. Hume and K. Johnson: Academic Press.
- Jacquemot, Charlotte, Emmanuel Dupoux, and Anne-Catherine Bachoud-Lévi  
2007 Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. *Cognitive Neuropsychology* 24 (1):3–22.
- Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene, and Emmanuel Dupoux  
2003 Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *Journal of Neuroscience* 23 (29):9541–9546.
- Johnson, Keith  
1997 Speech perception without speaker normalization. In *Talker variability in speech processing*, edited by K. Johnson and J. Mullennix, 145–166. San Diego: Academic Press.
- Jusczyk, Peter W.  
1997 *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Jusczyk, Peter W., Anne Cutler, and Nancy J. Redanz  
1993 Infants’ preference for the predominant stress patterns of English words. *Child Development* 64:675–687.
- Jusczyk, Peter W., Derek M. Houston, and Mary Newsome  
1999 The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39 (3/4):159–207.
- Jusczyk, Peter W., Paul A. Luce, and Jan Charles-Luce  
1994 Infants’ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33:630–645.

- Keating, Patricia  
 1990           Phonetic representations in a generative grammar. *Journal of Phonetics* 18:321–334.
- Kouider, Sid, and Emmanuel Dupoux  
 2005           Subliminal Speech Priming. *Psychological Science* 16 (8):617–625.
- Kouider, Sid, and Emmanuel Dupoux.  
 2009.           Episodic accessibility and morphological processing: Evidence from long-term auditory priming. *Acta Psychologica* 130:38–47.
- Kuhl, Patricia K.  
 2000           A new view of language acquisition. *Proceedings of the National Academy of Science USA* 97 (22):11850–11857.
- Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens, and Björn Lindblom  
 1992           Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255 (5044):606–608.
- Levelt, Willem J. M.  
 1989           *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, Willem J. M.  
 1992           Accessing words in speech production: Stages, processes and representations. *Cognition* 42:1–22.
- Lively, Scott E., John S. Logan, and David B. Pisoni  
 1993           Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal Of The Acoustical Society Of America* 94 (3 Pt 1):1242–1255.
- Lively, Scott E., David B. Pisoni, Reiko A. Yamada, Yoh'ichi Tohkura, and Tsuneo Yamada  
 1994           Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal Of The Acoustical Society Of America* 96 (4):2076–2087.
- Luce, Paul A., and Emily A. Lyons  
 1998           Specificity of memory representations for spoken words. *Memory & Cognition* 26 (4):708–715.
- Mattock, Karen, and Denis Burnham  
 2006           Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy* 10:241–265.
- Metz-Lutz, Marie-Noëlle, and E. Dahl  
 1984           Analysis of word comprehension in a case of pure word deafness. *Brain and Language* 23 (1):13–25.
- Morton, John  
 1969           The interaction of information in word recognition. *Psychological Review* 76:165–178.

- Morton, John  
1980 The logogen model and orthographic structure. In *Cognitive processes in spelling*, edited by U. Frith, 117–133. London: Academic Press.
- Myers, James  
1999 Lexical phonology and the lexicon. *Rutgers Optimality Archive* 330–699.
- Norris, Dennis, James M. McQueen, and Anne Cutler  
2000 Merging information in speech recognition: feedback is never necessary. *Behavioral and Brain Sciences* 23 (3):299–325.
- Pater, Joe, Christine L. Stager, and Janet F. Werker  
2004 The perceptual acquisition of phonological contrasts. *Language* 80 (3): 384–402.
- Pierrehumbert, Janet B.  
2001 Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency effects and the emergence of lexical structure*, edited by J. Bybee and P. Hopper, 137–157. Amsterdam: John Benjamins.
- Pierrehumbert, Janet, Mary Beckman, and Robert Ladd  
2000 Conceptual foundations of phonology as a laboratory science. In *Phonological Knowledge: Conceptual and Empirical Issues*, edited by N. Burton-Roberts, P. Carr and G. Docherty, 273–304. Oxford: Oxford University Press.
- Pisoni, David B., and Susannah V. Levi  
2007 Representations and representational specificity in speech perception and spoken word recognition. In *Oxford Handbook of Psycholinguistics*, edited by M. G. Gaskell, 3–18. Oxford: Oxford University Press.
- Prince, Alan, and Paul Smolensky  
1993 *Optimality theory: Constraint interaction in generative grammar*. New Brunswick: Rutgers University.
- Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport  
1996 Statistical learning by 8-month-old infants. *Science* 274:1926–1928.
- Sancier, Michele L., and Carol A. Fowler  
1997 Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25:421–436.
- Schriefers, Herbert, Antje S. Meyer, and Willem J. M. Levelt  
1990 Exploring the Time Course of Lexical Access in Language Production - Picture-Word Interference Studies. *Journal of Memory and Language* 29 (1):86–102.
- Sheldon, Amy, and Winifred Strange  
1982 The Acquisition of /r/ and /l/ by Japanese Learners of English: Evidence That Speech Production Can Precede Speech Perception. *Applied Psycholinguistics* 3 (3):243–261.

- Singh, Leher, Katherine S. White, and James L. Morgan.  
2008. Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development* 4 (2):157–178.
- Smith, Neil  
1973 *The Acquisition of Phonology. A Case Study*. Cambridge: Cambridge University Press.
- Stager, Christine L., and Janet F. Werker  
1997 Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388 (6640):381–382.
- Steriade, Donca  
1995 Underspecification and markedness. In *The handbook of phonological theory*, edited by J. Goldsmith, 114–174. Oxford: Blackwell.
- Swingley, Daniel, and Richard N. Aslin  
2000 Spoken word recognition and lexical representation in very young children. *Cognition* 76 (2):147–166.
- Swingley, Daniel, and Richard N. Aslin  
2002 Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science* 13 (5):480–484.
- Werker, Janet F., Leslie B. Cohen, Valerie L. Lloyd, Marianella Casasola, and Christine L. Stager  
1998 Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology* 34 (6):1289–1309.
- Werker, Janet F., and Richard C. Tees  
2005 Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology* 46 (3):233–251.