Draft. Not to be quoted without permission.

Do Experimental Studies Inform Theoretical Research?[1]

Madelyn Kissock
Concordia University

Abstract

The interpretation of experimental results and the subsequent determination of what implications, if any, those results have for theoretical models of language are complex tasks. This paper discusses these tasks in detail using two recent experimental studies, Maye, Werker, and Gerken (2002) and Werker, Fennell, Corcoran, and Stager (2002), to illustrate its points. The paper argues that while the common goal for both experimentalists and theoreticians is to understand the nature of language (here, specifically, acquisition phenomena), the dichotomy of goals and training between the two groups leads to a lack of clarity in exactly what aspects of linguistic knowledge are being tested and why. A direct result of this is that experimental studies can be, and frequently are, interpreted as support for a number of different theoretical claims even though detailed evaluation of the studies reveals little or no evidence for such support. The paper highlights areas in which a more explicit framework will allow a more fruitful interpretation of results and a correspondingly greater contribution to our understanding of language phenomena.

1 Introduction

We have come to accept the existence of a number of phenomena in the physical world for which we have no direct, experiential evidence (e.g., gravity, atoms, molecules, and so on). In spite of this, there is still the sense that arguments based on theory-internal assumptions and theories based on indirect evidence of phenomena alone are somewhat suspect. This feeling appears to be especially preva-

[1]The author wishes to thank the audience of the 10th Manchester Phonology Conference for helpful comments and suggestions. In addition, thanks to Mark Hale and Charles Reiss for their usual insightful comments.

lent among experimental researchers (perhaps not surprisingly, given their orientation and training). However, even those who work only within the most abstract, theoretical frameworks – who couldn't tell a test tube from an inner tube – would undoubtedly be happy to be able to say 'I told you so' if their theories were borne out by actual experimental results (or perhaps better yet, if a rival theory was disproven...). Research in the field of language acquisition is a good example of the dichotomy between the experimental and the theoretical researchers in both goals and training. The doubts of experimental researchers can be seen in statements such as the following from Maye and Gerken (2000: fn. 1) '...the psychological reality of the linguist's "phoneme," comprising multiple allophones, has not been experimentally demonstrated.' At the other end of the spectrum, there are any number of purely theoretical researchers who have little or no knowledge of the current state of experimental research in what might be considered the same general field.

While I acknowledge that gathering empirical support for theoretical postulates is a crucial part of scientific inquiry, I believe that the *type* of empirical support, available and necessary, in any particular case is going to depend critically upon the nature of the item for which that support will be offered. It is mistaken to believe that experimental evidence should be able to produce overt physical correlates of every item postulated as part of a theory. And, perhaps more seriously, to believe that elements which cannot be supported in this manner should be excluded from the theory. For those aspects of a theory which *are* amenable to experimental research, a clear and thorough evaluation of the results of experimental evidence is obviously crucial to the enterprise as a whole. This is just as true of results that might be considered 'negative' (i.e., instances where subjects fail to meet criterion) as results that are considered 'positive.'

The purpose of this paper is to highlight the complexity of interpreting experimental results and their subsequent implications for theoretical frameworks. Two recent experimental studies have been chosen as the focus – Maye, Werker, and Gerken (2002) and Werker, Fennell, Corcoran, and Stager (2002). The claim is not that these studies are fundamentally different from other experimental studies, but rather, that they combine a number of features which make them ideal for the purposes of this paper. First, they treat two topics concerning the 'initial state' of the phonological acquirer which are also being hotly debated in theoretical phonology circles. Second, although quite recent, the studies have received broad attention, are relatively widely cited and, therefore, are clearly having an impact. Third, the experimental researchers are, I believe, well-respected and experienced in their fields. What I will argue here is that only through the lens of an explicit theoretical linguistic framework can the results of experiments such as these be coherently interpreted.

The paper is organized as follows. Section 2 provides some general theoretical assumptions about the phonological component and its interaction with other components in language perception. Section 3 discusses two hypotheses about building featural representations from physical input and examines in detail experimental studies in support of one of the two hypotheses. In Section 4, I present briefly two positions concerning children's initial lexical representations and go on to discuss an experimental study which is related to this topic. Section 5 concludes with a summary of the paper.

2 General Theoretical Assumptions

There are a number of commonly held assumptions about adult phonologies which cut across theoretical frameworks. One of these is that phonological features, not 'speech sounds' or 'segments,' are the relevant primitives. These primitive symbols are used for phonological and phonetic representation and are manipulated by computational operations in the phonology. I assume, fairly uncontroversially, that the set of possible human language phonological features is innate – specified by UG. UG also specifies a set of possible computational operations, defined over the relevant features. In addition, I take the innate feature set to be finite and the actual number of features to be relatively limited, the large diversity of 'segments' in natural language being the result of the combinatorial possibilities of this feature set. A second fairly commonly held assumption is that, in the adult lexicon, phonological representations (often referred to as 'underlying' representations, abbreviated UR's) contain only idiosyncratic featural information whereas phonetic representations (which are the product of phonological computation and are sometimes called 'surface representations', abbreviated SR's) contain more featural detail.[2] Both the assumption regarding feature-based representation and that which distinguishes two levels of representation (for which I use the traditional 'UR' and 'SR' here) are based on a large body of evidence all of which is necessarily indirect – since features are not present in the physical data – but which is nonetheless compelling.

Interestingly, the measurable, physical data regarding sound produced in linguistic utterances suggests that virtually the opposite hypothesis should be en-

[2]The degree to which lexical representations may be underspecified is not immediately relevant to the topic of the paper.

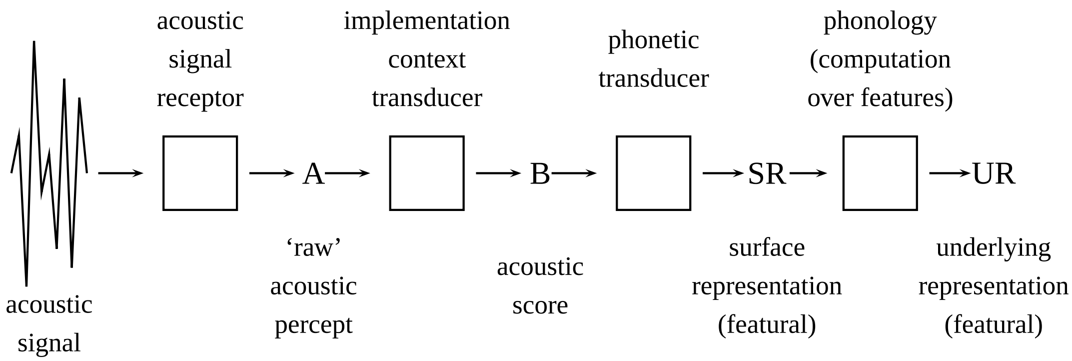
tertained – specifically, that there are *no* consistent representations such as features. Variation within the speech of an individual, cross-speaker variation, co-articulation, and other physical variables serve to produce physical evidence where identity between data samples is small enough to appear almost accidental. For example, ten different instances of the word *sit* uttered by the same speaker will vary at least slightly in the durations of the fricative, vowel, and stop closure, in amplitude, in the exact formant structure for the vowel, amount of noise for the fricative, and so on. If one compares the production of *it* by the same speaker, one will see that the vowel realizations of the [ɪ] of *sit* and the [ɪ] of *it* are dramatically different acoustically. The [ɪ] if *it* will have an abrupt onset and no transitional formants such as those present in consonant vowel co-articulation at the beginning of the [ɪ] of *sit*. Add a second person saying both of these words and these acoustic differences will be multiplied before even taking into account the differences due to differences in voice pitch (F0) of the individual speakers. Record both speakers as they stand on the top of Mt. Everest and then again as they stand in Bombay during the summer and one will see additional effects of air density upon each of their utterances. Crucially, in spite of all the actual physical variations in the utterances, a relevant pool of listeners will still correctly identify *sit* and *it*. This type of evidence in favor of abstract, feature-based representations is so strong that one is forced into the 'features as primitives' hypothesis, against all direct physical evidence. The question arises, then, of how adults reach a state where they extract consistent featural representations from such highly variable physical evidence.

The development of adult-state lexical representations is usually looked at as a two-part process. First, it involves the ability to map a physical input onto an abstract feature. This is a necessary precursor to the second part which is the ability to store a lexical entry with the appropriate (adult) featural representations. The experimental studies on infant speech perception discussed in this paper directly address these particular issues.

2.1 Perceptual Levels for Language

As mentioned earlier, I feel that it is crucial to have an explicit, and therefore falsifiable, framework within which experimental results may be weighed and interpreted. Because perception of speech signals involves both linguistic and extra-linguistic cognitive systems, it is necessary to make some assumption about the division of labor among these systems so that one does not, for example, improperly conflate acoustic-level processing with processing by the phonological component and vice versa. I will assume the model described in Hale, Kissock &

Reiss (to appear), here only illustrated for (adult-state) perception, not production.



The four broad categories distinguished in the figure are: 1) the raw physical signal and its reception (up to A); 2) components and processes specific to non-linguistic processing (from A to B); 3) components and processes specific to *language* (from B onward); and 4) components and processes specific to the grammar, in this case, the phonology (from SR onward). By reception of the raw physical signal, I intend the most superficial of processing – e.g., that of the middle and inner ear – the anatomical properties of this area determine the frequency range available to humans and physical anomalies in this area can result in hearing loss. By implementation context, I intend the myriad properties that *accompany* language-specific information on any particular occasion. These range across speaking rate and volume, information about the identity and emotional state of the speaker, and information about the external physical conditions which affected the signal at the time of utterance (air density, for example).[3] Stripping out these properties reduces the input to an 'acoustic score.' Information contained in the acoustic score is language-specific but not represented in *phonetic* features (i.e., not in a format that the grammar can manipulate).[4] The transduction process involved in converting the acoustic score to phonetic features 'normalizes' for co-articulation and other positional (e.g., word-initial, word-final) effects and produces a phonetic feature representation that the grammar can parse. I assume that transduction processes are mechanical in the sense that they systematically convert a given input to a given output. As the figure indicates, I take the phonological

[3]These are compressed into a single box only for convenience here. I assume that the actual process is extremely complex and requires interaction with a number of other cognitive areas.

[4]Since it seems likely that a number of other, if not all other, cognitive processing mechanisms use featural representations of some sort, I emphasize 'phonetic' to distinguish those specific to the grammar.

component of the grammar to include only mappings of featural representations to featural representations (SR to UR here), following Keating (1988).

2.2 Where Differing Input May Fail to be Distinguished and Why

In the above model, there are four levels at which two physically different inputs may fail to be distinguished from one another, resulting in two physically different stimuli being judged as 'same' instead of 'different.' Two instances of *sin*, one produced by an adult male (sin-M) and one by an adult female (sin-F) will illustrate how judgements of 'same' or 'different' might be made at particular levels. The first level is the initial point of signal reception. Athough obviously critical to any further processing, I assume that all language stimuli provided are within the human capacity frequency range and that, in the cases under discussion, hearing is normal. Therefore, sin-M and sin-F will be judged 'different' at this level assuming, as I am, that the pitch of the male's voice is different than the pitch of the female's voice. Notice that beyond this initial level, there is considerable evidence that the linguistic and non-linguistic components of the system operate quite independently of one another. Here the judgements about 'same' or 'different' will depend crucially on the task given to the listener. On one hand, in word-matching tasks, a listener who hears sin-M and sin-F will judge the two tokens to be the same – both are *sin*. If, on the other hand, the listener is queried about the *source* of the words (same or different) they will report 'different.' This judgement of 'different' will be based purely on some voice recognition mechanism which attends to frequency cues in order to identify particular speakers, or simply male vs. female speakers, as in this instance.[5] Crucially, for present purposes, the adult ability to map multiple physical inputs to a single feature bundle does not mean that adults will fail to discriminate between the physical inputs at some other level. A listener's judgement of 'different' based on some non-linguistic property such as voice pitch is what I take to be the second level for judging inputs. At purely grammatical levels (third and fourth levels), a listener might fail to respond to a difference if one or more of the features in the input representation from the phonetic transducer is not able to be matched within the grammar itself at either SR or UR. This inability to match a feature would result in certain non-native feature bundles being conflated with native feature bundles by the grammar – failure of native English speakers to respond to the retroflex/dental contrast, for example

[5]Other types of source information may be present, of course, such as location of the speakers.

(cf. Werker and Tees, 1984).[6] Finally, one would expect a difference in judgement based on whether the features present were predictable in their distribution within a particular grammar or idiosyncratic. Predictable features, such as nasalization on the vowel of *sin*, should be present and be able to be judged at the SR level (e.g., presence or absence of nasality on a vowel) whereas the same predictable feature (nasality) should be absent at UR, resulting in a judgement of 'same' for the nasal/non-nasal contrast. Note that this failure to respond to a contrast has a different cause than the previous failure. Here, the features (and feature bundles) are clearly matchable by the grammar at *some* level, just not at every level.

3 Mapping Signals to Features

There are competing theories regarding the acquisition of mappings between physical signals and phonetic features. The two under discussion here broadly represent the two logical possibilities, either 1) the mappings are innate (meaning that the adult ability to convert a variety of physical signals to a single featural representation is present at birth); or 2) mappings are acquired only through experience with the physical data itself (specifically, the features are innate, but mapping signal to features is experience-driven). The second of these is especially appealing in light of what appears to be language-specific variation for sounds that phonologists have taken historically to be featurally non-distinct – the difference in height of Danish [i] vs. English [i], for example. Recently, many researchers who support experience-driven mapping have posited statistical learning algorithms of various types.[7] The experimental study that I will examine in detail in this section is based on such an algorithm. While my own position is that experience-driven learning cannot result in adult-state knowledge (and therefore the knowledge must be innate), the purpose here is simply to explore the support of this particular experimental study for the claim about statistical learning, not to adduce support for the innateness position.

Although the focus is on acquisition, the hypothesis that infants can use statistical information to build signal-to-feature representations is based on the assumption that humans, in general, are able to process and use information in this way. As background for the infant study, then, I first present an experimental study on

[6] See Hale and Kissock (1997) for more discussion.

[7] A number of other models exist as well, some based upon learning motor control patterns (Kingston and Diehl, 1994) and others based upon acoustic 'space' in some way (Kuhl, 1991; Stevens,1972; Liljencrants and Lindblom, 1972).

adults by Maye and Gerken (2000), henceforth MG. This study was chosen over others because the stimuli appear to be identical to the infant study and it seems apparent from this and other aspects of the MG study that the infant study by Maye, Werker, and Gerken (2002), henceforth MWG, was the natural follow-up to the MG study.

3.1 Background – Maye & Gerken 2000 (MG)

MG favor a distribution-based learning model, specifically, they "...argue against the hypothesis that phonemes are learned via a comparison of minimal pairs, and provide evidence in favor of the hypothesis that phonemes are learned from the distribution of sounds in a language." (MG:1). The title of the paper is, correspondingly, 'Learning Phonemes Without Minimal Pairs.' As it happens, the MG title and virtually all subsequent discussion of what the experiment will or has shown are quite misleading, at least when viewed from a linguistic framework. This problem appears to stem from a fundamental miscommunication or lack of communication between the fields of psychology and linguistics.[8] How the confusion arises is illustrated in the first footnote of the paper (which goes with the first instance of the term 'phoneme' in the first sentence of the Introduction), repeated below:

> 1. The categories we would like to account for are not "phonemes," as the term is used by linguists. For linguists, the term applies to categories that include multiple allophones, which appear in different phonological positions. What we are interested in is perhaps more appropriately termed "phonetic categories" or "phonetic equivalence classes;" specifically, sounds which are categorized together in a *particular* phonological position. However, we point out that the psychological reality of the linguist's "phoneme," comprising multiple allophones, has not been experimentally demonstrated. The categories we account for here could plausibly be the *only* psychological correlate to phoneme categories. (MG:2)

As this footnote specifies, MG are not actually talking about phonemic categorization but, instead, phonetic categorization. Since the 'minimal-pair based learning'

[8]This is probably not helped by the fact that linguists fail to be explicit about such terms as 'phonetic' and use equally ambiguous notational devices. For example, square 'phonetic' brackets are used by linguists to indicate everything from the abstract featural representation that is the output of the grammar to the physical signal itself.

referred to in the title of MG is posited only for truly phonemic categorization, it is not clear what hypothesis MG are ultimately arguing against. MG proceed to point out that a number of studies have shown that there is a change in discrimination response patterns for infants once they reach a certain age bracket (earlier for vowels than for consonants), specifically, that infants fail to respond to contrasts which they did respond to at earlier ages.[9] Since the age brackets for this change range from 6 months (cf. for example, Kuhl 1991, for vowels) to 10-12 months (cf. for example, Werker and Tees 1984, for consonants), it is extremely unlikely, again as MG point out, that this change in response is due in any way to the acquistion of a lexicon, there being no evidence that infants have a lexicon at these early stages.[10] I deduce, then, that the purpose of MG's study is to provide evidence for a distribution-based algorithm to account for this change in response which seems to be due to the absence of certain sounds in the input from the surrounding environment.

I take this, ultimately, to be a potential account for how a variety of physical signals are collapsed into a single feature bundle. The reasoning is as follows. One can fairly safely rule out some type of actual hearing loss – subjects can be and are tested for this.[11] In addition, in the model that is proposed here, transduction is mechanical – the same token will always be transduced in the same way. This leaves us with only the grammar as the locus for reduced sensitivity. From the standpoint of the grammar, the tokens used in the relevant discrimination tasks can differ from one another only *featurally*, whether at the level of SR or UR. So, for example, phonological computations which are sensitive to the features of a simple coronal [t] will not be triggered by the features of a retroflex [ʈ]. The reduction in sensitivity must, then, be some phenomenon that is related to features.

In fact, I believe that MG (and others) are approaching a much larger question than just that of why and how infants begin to behave differently on discrimination tasks involving non-native sounds after a certain amount of exposure to the environment language (in itself a major question). The issue of the 'many-to-one' signal-to-feature reduction process is relevant to tokens within the environment language as well – tokens that count as the 'same' for phonological purposes (and

[9] The contrasts in question involve tokens of sounds, one of which is present in the environment language and one of which is absent. Occasionally the non-native token is actually present as a non-contrastive variant, however (cf. Trehub, 1976, for nasalization on vowels).

[10] Not all non-native contrasts are lost at these early ages. See Burnham (1986) for more discussion of this.

[11] In addition, studies suggest that adults will *hear* certain contrasts if the interstimulus interval is shortened. This is discussed further in section 3.3.

therefore must have the same featural representations), but are physically different from one another. In addition, there is the cross-linguistic variation in the physical 'target' spaces for what, featurally, are thought to be the same representations. This variation takes several forms: 1) the target space may be non-equivalent, non-subset but intersecting, as in the case of Danish [e] and Japanese [e]; and 2) the target space may be broader for one speaker than for another speaker, with the difference standing in seeming correlation with a difference in the relevant segmental inventory. For example, one target space may be a proper subset of the other, as in the case of a typical 3 vowel system, with a relatively 'broad' target space for [i] (for example), vs. a system with a larger vowel inventory and correspondingly reduced target space for [i].[12] If a statistical learning algorithm is viable for the environmental language effects, then it seems likely that the same algorithm could be responsible for these other phenomena.

MG propose a distribution-based learning algorithm which is crucially dependent upon the existence of a bimodal or unimodal distribution of tokens. Again, the overall purpose of their study was to determine whether humans were sensitive to such distributions and would categorize tokens into groups based on those distributions. The hypothesis is that exposure to a bimodal distribution will cause subjects to posit two distinct (phonetic) categories whereas exposure to a unimodal distribution will result in only a single category. Because the hypotheses and stimuli were, as far as I can tell, identical for the MG study with adults and the MWG study with infants, I will only discuss one of the two studies in detail – the MWG study. I chose MWG because their ultimate goal is to explain an acquisition phenomenon (which leads to the adult state), among other reasons. For the moment, I note only that MG concluded that "The results of this experiment support a distribution-based model of phoneme [read 'phonetic category' MJK] learning." (MG:1)

3.2 Maye, Werker and Gerken 2002 (MWG)

MWG's basic assumptions are that 1) infants have a way of compiling statistical information from the input (building on studies such as the MG one, which conclude that adults have this ability); and 2) a bimodal distribution of hits (from the physical input) will result in the child positing two distinct phonetic categories, whereas a unimodal distribution will result in the child positing a single

[12] For a detailed discussion of these sorts of variations as well as hypotheses regarding their sources, see Hale, Kissock & Reiss (to appear).

phonetic category. Note that, in this study, the authors use the term 'phonetic category,' which, while still somewhat amibiguous, avoids the 'phoneme' problem discussed for MG. Various forms of 'perceive' are typically used when discussing infant sensitivities' to tokens. 'Perception,' unlike 'phonetic,' covers the entire range of possibilities. It may be used to mean 'hear' in the most basic sense of processing acoustic signals at the ear level. It may refer to non-linguistic differences in source of the signal, for instance, speaker identification. Equally, it may refer to linguistic levels such as SR and UR. For reasons cited above, I assume that actual hearing ability is not one of the options being considered for the interpretation of 'perception' here. Apart from that, the intended perceptual level is not entirely clear. MWG predict that:

> If a given acoustic property is non-contrastive, however (that is, it does not differentiate between two categories), speech sound tokens will fall into a single (potentially wider) cluster, forming a unimodal distribution. (MWG: B103)

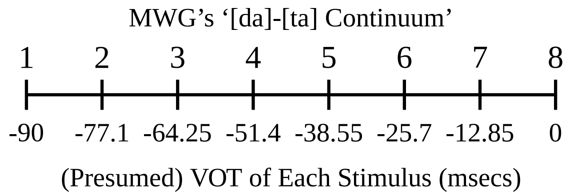
Since the categories that MWG are referring to are cited elsewhere as 'phonetic categories,' I take 'acoustic property' above to mean some physical correlate of a phonetic feature (voicing, for example). The methods and results of the experiment are summarized in the following section.

3.2.1 Synopsis of the experiment and results

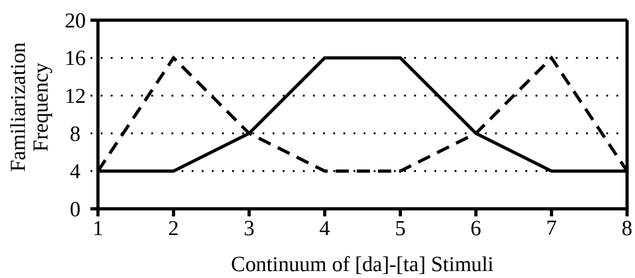
The participants in the MWG study were English-environment infants, 6-8 months old. The test was a type of preferential looking procedure with alternating and non-alternating trials after an initial familiarization phase. The authors state that '...a novelty preference would be evident in longer looking times on Non-Alternating trials. However, any significant difference in looking times for the two types of test trials would indicate discrimination of the test stimuli.'(MWG: B106) The stimuli were tokens of [da] and [ta] (unaspirated [t]) which varied along an

> eight-point continuum from [da] to [ta]...The phonetic difference between [da] and [ta] was in the presence of prevoicing (VOT $\geq$ -90 ms) for [da], as well as in the trajectories of the first two formants, from vowel onset to vowel center. (MWG: B106)

Below, I show the interpretation of the above description of token variation by VOT in graphic form. Zero represents, as is usual, release of the stop closure.



The familiarization phase included the same total number of tokens for each subject, but one test group had more tokens from the extreme ends of the continuum (bimodal distribution) and the other test group had more tokens from the central range of the continuum (unimodal distribution). In addition, two filler syllables [ma] and [la] were included in the familiarization phase. MWG's figure showing the distributions of tokens during the familiarization phase for each group is reproduced below, along with its label (MWG: B104). Frequency is plotted on the y-axis and the eight-point continuum is plotted on the x-axis. The Bimodal group is designated by the dashed line and the Unimodal by the solid line.[13]



The test phase itself consisted of two types of trials, alternating and non-alternating. The alternating trials contained tokens 1 and 8 from the continuum while the non-alternating trials contained either only token 3's or only token 6's. The hypothesis about distribution-based learning predicts that, at the end of the familiarization phase, the Bimodal group will have developed two distinct phonetic categories because of the bimodal distribution pattern of the stimuli. The Unimodal group, on the other hand, is predicted to collapse the stimuli into a single phonetic category based on the (relatively central) distribution of stimuli from the

[13]In fact, the numbers on this graph do not correspond to the prose description of the number of tokens during the familiarization phase. MWG state that "...infants heard six blocks of 24 syllables each. Each block was composed of 16 stimuli from the [da]-[ta] continuum...and the eight filler syllables." (MWG: B106) The graph indicates that each infant heard 64 total tokens from the [da]-[ta] continuum whereas the prose indicates that the infants heard 96 tokens (6x16).

familiarization phase. Crucially, further tokens from the [da]-[ta] continuum presented during the test phase would then be perceived differently by the Bimodal group (who had two categories to choose from) than by the Unimodal group (who had only one category). The Bimodal group would be predicted to notice the contrast on the Alternating test trials and the Unimodal group would be predicted *not* to notice that contrast. For the Bimodal group, the most 'novel' types of test trials would be those that did *not* alternate (because they were accustomed to hearing alternating stimuli).

The result of these test trials was that infants who received the bimodal distribution (greater numbers of tokens from what MWG called 'endpoint' regions) during the familiarization phase looked longer on Non-Alternating than on Alternating test trials. And MWG state also that

> ...infants from the Unimodal condition showed no preference, indicating that only infants in the Bimodal condition discriminated the test stimuli. Because the infants were familiarized to strings of varied syllables, the Bimodal infants' longer looking times on the Non-alternating trials exemplify a novelty preference. (MWG: B108)

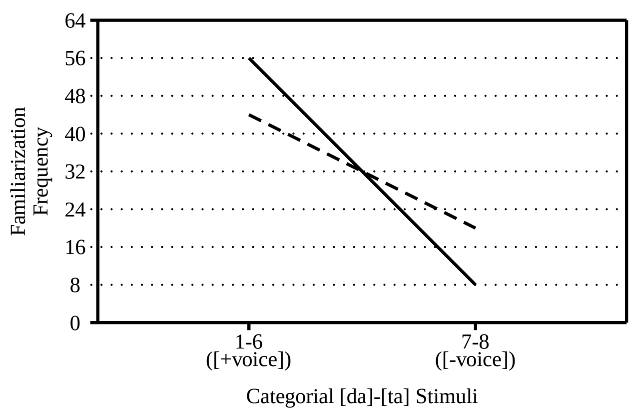
The overall conclusion of the authors was that this study supported the claim that infants can (and, presumably, do) use a distribution-based learning mechanism to build phonetic categories.

3.3 Discussion

The MWG study, in my opinion, suffers from several rather serious problems. The first of these is the choice of stimuli, the second is the perceptual level or levels that the experiment taps into, and the third is the expectations with regard to the novelty preference. Each of these will be discussed in turn in the sections below.

3.3.1 Choice of Stimuli

The distribution-based learning algorithm requires that the stimuli be able to range over a continuum such that some stimuli can be considered more 'central' (unimodal) and some more 'extreme' (bimodal) with respect to the end points of that continuum. Perception of the voicing distinction on stops has been rather convincingly shown to be *categorical* – i.e., it is exactly *not* a continuum (8-point or otherwise). Moreover, phonetic feature representations correspond to these (discrete) categories (again, based on evidence from phonological processes). The categorical nature of this perception leads both adult and infant subjects to show poor to no discrimination of within category differences in VOT and fairly sharp boundaries between categories (Lisker and Abramson 1970, Eimas et al. 1971, inter alia). Given categorical perception, there was no bimodal distribution from a *linguistic perceptual* standpoint for the participants (only in the graph showing the *acoustic* output of the Bose speaker). On one estimate, tokens 1-6 (-90 to -25.7) would have been indistinguishable from one another (assuming that one of the VOT category boundaries lies somewhere in the vicinity of -20 ms.). The resulting distribution during the familiarization phase would then resemble that illustrated in the figure below where the dashed line once again represents the 'Bimodal' group and the solid the 'Unimodal' group.



Note that, from the perspective of the participants, there would be no bimodal/unimodal distinction between the two groups' familiarization phases.

One possible explanation for considering the stimuli in this study to be a continuum is, of course, that MWG do not adopt the position that there is categorial perception.[14] In that case, the study would contribute to a body of literature which has argued against categorial perception (cf. Harnad, 1987). However, MWG state clearly in their abstract that it is developmental *changes* in speech perception due to native language input that they are interested in explaining. This implies that there are initial-state speech perception abilities (e.g., categorical perception of VOT) and that these initial-state abilities are *not* under discussion in their paper.

Another possible explanation is that MWG are conflating the types of perception that they are tapping into – providing stimuli which can be processed at the

[14]Categorial perception has been widely discussed in the literature for many years. While the presence or absence of such perception is still debated, support for categorial perception seems sufficiently widespread that the term itself deserves mention by the authors if that is, in fact, one of the things that is being tested in their study.

level of acoustic signal reception but interpreting the results as if they were necessarily processed by some level of linguistic perception. The issue of levels of perception and/or processing is complex and crucially related to the interpretation of experimental results. I discuss these matters in detail below.

3.3.2 Perceptual level-related differences

Werker (1994) summarizes the results of a series of sophisticated and interesting experiments where positive identification of contrasts between stimuli correlated with interstimulus interval (ISI) time (Werker and Tees 1984; Werker and Logan 1985). Briefly, subjects who failed to respond to non-native contrasts at a 1500 msec ISI did respond when the ISI was shortened to 500 and 250 msec. Werker (1994: 130) states:

> In an attempt to make sense out of this pattern of findings, we proposed that subjects can use one of three different processing strategies–phonemic, phonetic, and acoustic–depending on the interstimulus interval. When tested with an ISI over 500 msec, subjects appeared to use a phonemic processing strategy and were unable to discriminate the nonnative contrast. Thus, when the ISI is long, subjects seem unable to discriminate two stimuli unless they can assign them distinct linguistic labels. At shorter ISIs, subjects showed evidence of using both a phonetic and an acoustic strategy. Evidence for a phonetic strategy was provided by subjects who could discriminate retroflex from dental exemplars but could not discriminate among the several exemplars within either phonetic category. Evidence of acoustic processing was provided by subjects who could discriminate between the several retroflex or the several dental exemplars. These findings indicate that adult listeners can discriminate between tokens on the basis of phonetic and acoustic information *if the task requires it* [emphasis MJK] but that the most readily available strategy is to perceive speech stimuli in terms of native-language phonemic categories.

This seems like the only logical explanation of the results of such experiments, given any model which distinguishes between the grammar, which uses only (phonetic) featural representations, and extra-grammatical processing.[15]

[15] In the case of the retroflex/dental contrast mentioned, it is not completely clear what the status of a dental was for the native English speaking subjects. Dentals are typically considered surface variants of coronals, with no phonemic status of their own.

The implications of such findings are far-reaching. Any experimental studies on adults involving sound contrasts could, conceivably, be tapping into one of three points at which subjects can make a ‘same’ or ‘different’ judgement. (In this case, the non-linguistic, non-acoustic level for speaker recognition and the like is left aside.) While in the case discussed in the above quote the level tapped into seemed to be a function of ISI, it would not be surprising if the levels queried were also sensitive to 1) other task-related effects; and/or 2) the demand of other attentional resources that it may be difficult to control for; and/or 3) the nature of the signal itself. In the case of infant speech perception studies, the issue becomes even more complex. While it is true that under a certain age, the phonemic level is not present, since its existence depends upon the existence of a lexicon, the ability to determine whether tokens are being processed at the phonetic or acoustic level is likely to be more difficult.

There is no reason to assume that VOT and categorial perception should have any different status than the retroflex/dental contrast discussed by Werker (1994). One predicts that small, within-category, differences in VOT should be able to be distinguished at non-linguistic levels with a shorter ISI whereas the same small, within-category differences would not be distinguished at a linguistic level with a longer ISI.[16] It appears that the MWG study did not control for this factor, for some reason. The ISI during the familiarization phase was 500 msec. but the ISI during the test phase was 1 second.[17] The ISI of 500 msec. for the familiarization phase in MWG’s study is not predicted to tap into anything more than a non-linguistic level of processing. However, I take the establishment of phonetic categories to be a *linguistic* task which would assign featural representations to multiple inputs. It is unclear that the infants had adequate time to do this, given the ISI. In any event, variation in the ISI from the familiarization to the test phase may mean that the two phases tested different levels of perception and cannot, therefore, be combined for the purpose at hand.

[16] Pisoni, Lively, and Logan (1994: 130-1) note regarding VOT that “When the experimental conditions are modified to reduce uncertainty or when the subjects’ attention is explicitly directed to acoustic, rather than phonetic differences between stimuli, subjects can accurately discriminate very small changes in VOT (see also Carney, Widin, and Viemeister 1977).”

[17] In the MG study, on the other hand, the three phases – practice, acquisition, and test – also had different ISI’s. The practice phase ISI was 500 msec., the acquisition phase ISI was 1 second, and the test phase ISI was 500 msec. The acquisition phase corresponds to the familiarization phase of MWG and so the ISI’s of the corresponding phases in the two studies are exactly reversed.

3.3.3 Novelty preferences and the stimuli

There is an additional difficulty, in my opinion, in the prediction regarding which groups should show a novelty preference. The authors state that the Bimodal group's longer looking times on non-alternating trials indicated a novelty preference (because they had been familiarized to *alternating* stimuli). Specifically, the Bimodal group had heard a greater number of sounds from the extreme ranges of the continuum, i.e., more tokens which were more dissimilar acoustically, than the Unimodal group. As mentioned earlier, this would, by hypothesis, have led the Bimodal group to form two separate phonetic categories. The Unimodal group, on the other hand, heard greater numbers of tokens which differed little from one another – they all clustered in the same relatively narrow acoustic space. Based on this grouping of more nearly identical tokens, the Unimodal group was predicted to form only a single phonetic category (and would, presumably, be supposed to count hits from *any* point in the continuum as belonging to that category).

However, if human infants really show a novelty preference, what does one actually predict will be the behavior of the Unimodal group? Note that *all* infants (Unimodal and Bimodal groups) heard varied syllables in the familiarization phase because the syllables [ma] and [la] were *always* present as filler syllables. "...we included four tokens each of filler syllables [ma] and [la] during familiarization, in addition to the eight [da]-[ta] stimuli." (MWG: B105) The *Unimodal* group, therefore, should have looked longer at the Non-Alternating trials, too. *All* infants had a varied familiarization phase, so if a novelty preference exists, *all* infants should have looked longer at the non-alternating trials. The claim cannot be that a novelty preference exists for only some infants and those infants happen to fall into the Bimodal group. Nor can the claim be that only alternations between [da] and [ta]-type tokens and not [ma] and [la]-type tokens count as alternations.

A more general question arises with the distribution-based learning studies of both MG and MWG and, indeed, with all experimental studies on perception that depend upon some type of 'training' (or, perhaps more accurately, 'retraining') during a familiarization phase. In the MWG study, the English environment infants had 6 to 8 months of exposure to stimuli in their environment (in the MG study, 18-41 *years*). It is difficult to know how to judge exactly what countereffect 2.3 minutes of familiarization time (in the MWG study) has against 6-8 months of prior experience.

3.4 Interpreting the Results

All of the above criticisms notwithstanding, it is still true that the responses of the Bimodal group in the MWG study were different from those of the Unimodal group. I repeat MWG's Table 1 listing mean looking times for the groups below (MWG: B107).

Table 1
Mean (SE) looking times for infants in each age group and familiarization condition on Alternating and Non-Alternating Trials

	Alternating trial (s)	Non-Alternating trial (s)
6 months Unimodal	4.85 (0.47)	4.53 (0.51)
8 months Unimodal	4.98 (0.63)	5.20 (0.56)
6 months Bimodal	5.66 (0.44)	6.41 (0.32)
8 months Bimodal	5.45 (0.52)	6.15 (0.56)

The authors point out two aspects of the results which are summarized in the Table. First, that age (6 vs. 8 months) seemed to have no appreciable affect on performance. Second, that the Bimodal groups "...looked longer on Non-Alternating test trials than on Alternating trials, while infants from the Unimodal condition showed no preference, indicating that only infants in the Bimodal condition discriminated the test stimuli." (MWG: B108) What the authors did not point out specifically (other than to list it with the rest of the results in the Table) was that the Bimodal groups looked longer *at everything* than the Unimodal groups. I am no more able to account for this last feature of the results than for the features highlighted in the quote, but I also do not find the authors' interpretations convincing. The issues of catgorical perception, perceptual levels tapped into, and which groups are predicted to show novelty preferences all conspire to cast a wall of confusion around the interpretation of this study's results.

4 Lexical Representations

4.1 Theoretical Positions

The majority (and traditionally-held) opinion is that children's initial phonological representations are impoverished – they do not contain as complex a degree of featural specification as adult representations. A variety of instantiations of this position exist (Jakobson, 1941; Rice & Avery, 1995, *inter alia*). Evidence taken to support this position comes from a number of sources – production data (children fail to hit adult-like targets, 'simplify' consonant clusters, produce unusual consonant harmony patterns, and the like); theories of markedness and the relationship to initial ranking of constraints in an OT grammar, and others. Since a considerable body of evidence from comprehension data seems to indicate that children are, in fact, sensitive to fine featural distinctions (both at the level of sound contrast and word contrast), a number of researchers are forced to posit either a production grammar distinct from a comprehension grammar (Kiparsky and Menn, 1977) or two separate lexicons (see Menn and Mathei (1992) and citations therein) or else some additional mechanism which links the underspecified production representations to fully-specified comprehension representations.

Positions on the specificity of initial featural representations are directly related to two theoretical issues: 1) the initial constraint ranking in an OT grammar (the relative positions of $\mathcal{F}$-constraints vs. $\mathcal{M}$-constraints); and 2) general theories of markedness. Time constraints, the complexity of the issues, and the fact that they are somewhat orthogonal to the main topic preclude further discussion of them here. (For a discussion of production and comprehension within an OT framework, see Smolensky, 1996.)

The minority opinion holds that children's initial phonological representations are fully-specified. Although not directly concerned with initial lexical representations, the general arguments in Yip (1996) and Inkelas (1994) support this position. Yip (1996:30) states:

> There is a residue of evidence for abstract UR's, but the abstraction in one case (nasalization) involves an absence of localization in UR, not an absence of melody, and the other case, poetic rhyme, is inconsistent. One suspects that finding conclusive arguments for any particular UR may prove elusive. *This raises the possibility that it is also a difficult acquisition task, and that the easiest strategy may be to stick close to the phonetic surface form absent clear guidance to the contrary.* [emph. MJK]

An argument for underspecification being due *only* to optimization across the lexicon (and not, therefore, to a state of underspecified *initial* representations) is presented in Inkelas (1994:1).

> underlying representation is determined solely by optimization with respect to the grammar, not by imposing any type of constraints on

underlying representation ... [this] results in the use of underspecification only when there are alternant surface forms...

Hale and Reiss (2002), on the other hand, focus directly on acquisition and give arguments in favor of fully-specified initial representations. While an evaluation of the theoretical claims regarding initial representations is not the goal of this paper, since it is the less familiar as well as the less popular view, one argument for this position is presented briefly below.

4.2 The Subset Principle Argument (Hale & Reiss 2002)

The Subset Principle argument has two components: 1) the type of evidence available to the acquirer; and 2) a learning path which must proceed from the most specific to the least specific. More particularly, if an acquirer has an initial featural representation whose realization space is a superset of that of a particular adult representation, there is no evidence that will allow the acquirer to arrive at the adult representation, since all 'hits' in the adult space will be consistent with the acquirer's current representation. That is, the only learning path which will successfully lead the acquirer to a *final state* in which some representations are underspecified (and some are not) is the path from a subset (full specification) to a superset (underspecification). Because there tends to be considerable confusion about how this is instantiated in the domain of phonology, I will use the classic syntactic example as an illustration of the general point first.

In English, anaphors must have an antecedent within a relatively narrow syntactic domain, as the co-indexation subscripts in the example sentence below indicate.[18]

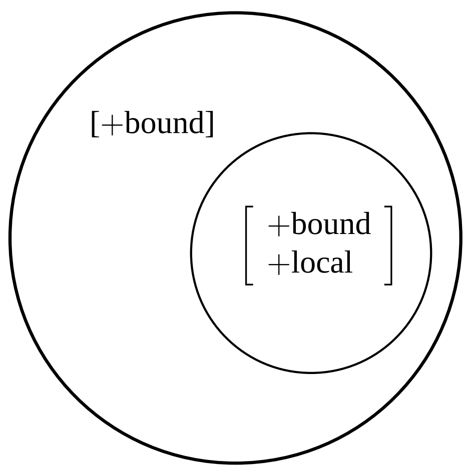
$$\text{John}_i \text{ asked Bill}_j \text{ to shave himself}_{*i/j}.$$

In Icelandic, on the other hand, it is possible to have an antecedent within the larger domain of the sentence, as opposed to only a clause-mate antecedent. (This is illustrated with an approximation of the corresponding sentence in Icelandic.)

$$\text{John}_i \text{ asked Bill}_j \text{ to shave self}_{i/j}.$$

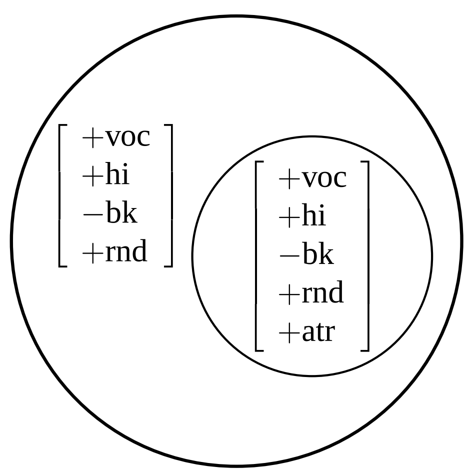
[18] Only those conditions on anaphor binding which are directly relevant to illustrating the subset principle are discussed here.

The distinction between the English and Icelandic anaphors is assumed to be determined by features present on the anaphors. Since the domain in which an antecedent can be found is *narrower* for the English anaphor, it will require more features on the anaphor to specify that domain. For example, if [+ local] is a feature that restricts the anaphor antecedent relationship to a clause, then the features of the English anaphor would be [+ bound] (a property of anaphors in general) and [+local] (a property of only some anaphors). The Icelandic anaphors, since they are not restricted to a clause-mate antecedency, will just have the feature [+ bound]. In terms of subsets and supersets, a more specific realization space (= a narrower domain) is a subset of the less specific realization space (= a broader domain). Crucially, *more* features are required to specify the subset realization space than are needed to specify the superset realization space. The set relationship between the English and Icelandic anaphors is illustrated graphically below.



The set of English anaphors is contained within the set of Icelandic anaphors – the relevant entities being the *domains* not the number of features required to specify such domains. From an acquisition perspective, the first hypothesis that the child makes about the features of the anaphors that he/she is exposed to must be that they correspond to the *English* set – the subset, or more restricted set. Only this initial hypothesis will produce final states of *both* the English and Icelandic type. If the acquirer posits [+ bound] and [+ local] as the feature set for anaphors first and is exposed to English-type anaphors, then they can just maintain that initial hypothesis. If, instead, the acquirer is exposed to Icelandic-type anaphors, they can reduce the feature set for anaphors to simply [+ bound]. Adopting a position which reverses this order of feature bundle acquisition such that the child's first hypothesis is that anaphors only have the feature [+ bound] will allow one to account for the acquisition of Icelandic-type anaphors but not that of English-type anaphors. Failure to acquire the English anaphor type under this latter scenario is due to the fact that there will be no evidence in an English environment which will be inconsistent with the child's initial hypothesis. (This is the 'no negative evidence' problem.) The child needs some type of evidence that the feature or features they have posited are incorrect in order to change the feature set along some dimension. If a child in an English environment posits only [+ bound] for anaphors, no amount of additional data will conflict with this hypothesis (all English anaphors are, indeed, bound) so there will be no reason for the child to adjust the features for anaphors and the child will not acquire the English anaphor feature set (and will, therefore, maintain a larger domain for anaphor-antecedent relations than is appropriate for English).

For phonology, the relevant comparanda will be realization spaces (= domains within which all hits are related to a single feature bundle) for phonetic feature bundles. Here, too, the number of features for any particular domain representation is inversely correlated with the range of compatible realizations for that representation. The figure below represents the relationship between a superset vowel space (which produces hits throughout *both* the [i] and the [ɪ] space) and one subset of that – a representation which produces hits only in the [i] space (represented by the features of the smaller circle). It seems counterintuitive just looking at features that the subset relations are being accurately represented in this graphic, but recall that it is not the features that are being compared, it is the realization spaces that these feature bundles produce.



I will approach the learning path from the majority position first. An initial impoverished representation for [sæd], for example, may contain only features that distinguish consonants from vowels, represented as:

/[+cons][+vocalic][+cons]/ = initial representation of [sæd]

At some later point, the grammar becomes able to parse an additional feature, say [+voice]. The acquirer then reviews the UR's of all his/her stored lexical items to see if any of them are now *inconsistent* in light of new feature. (Inconsistency leads to revising the features of representations, as in the anaphor case.) When the acquirer asks 'Is there anything wrong with my representation of '[+cons][+vocalic][+cons]' (aka [sæd])?, the answer will be 'No!'. There is nothing about having access to the feature [voice] which is inconsistent with having the representation '[+cons][+vocalic][+cons]'. As a result, the acquirer will not alter the initial representation for [sæd] (and will not attain an adult-like representation). The positive evidence that an acquirer gets during acquisition will only lead to changes in representations if there is a *conflict* with an existing representation. In the case of fully-specified initial representations, there is, first of all, no point at which more features become available – all features are present and accessible for forming feature bundles from the start. Developing an adult-like representation for [sæd] will entail a different process, lexicon optimization, which will reduce the UR to the minimally contrastive feature bundles. Note that lexicon optimization is not only a separate process but is necessary under the impoverished initial representation hypothesis, as well.

4.3 Previous Experimental Evidence

Obviously, the fully-specified initial representation hypothesis is dependent upon the pre-existence of fully-formed percept-to-feature mapping by the time the first lexical item is acquired. I have discussed the difficulties involved in distinguishing between simple acoustic perception and linguistic (i.e., featural) perception by infants in Section 3. While a number of experimental studies have been done which indicate that infants are sensitive to contrasting sounds of natural language, that same reaction is predicted if the infants are only using acoustic information, not featural information (cf. [b]-[p] Eimas et al. 1971 (1-4 mo. English infants); [a]-[ã] and [z]-[ř], Trehub 1976 (5-17 wk. Anglophone Canadian infants); [r]-[l], Tsushima et al. 1994 (6-8 mo. Japanese infants); [ʈ]-[t̪] and [tʰ]-[dʰ], Werker et al. 1981 (6-8 mo. Hindi & English infants); [ʏ]-[ʊ], Polka & Werker 1994 (4 mo. English infants); [ʇ]-[ʖ], Best, McRoberts & Sithole 1988 (6-8 mo. English infants); [k']-[q'] and [|]-[‖], Best et al. 1995 (6-8 mo. English infants)).[19] There is *some* evidence to suggest that processing goes beyond the acoustic level in studies that test for categorical perception with relatively small changes in VOT across the stimuli but where infants fail to respond to within-category token differences. In addition, studies which report on changes in response based on environment language (Werker et al, 1981, inter alia multa) support the notion that contrasts are being made (or failing to be made) based on something specifically linguistic (which I take to be featural representations). Again, these studies only suggest that infants *can* make judgements at a featural level, not that they necessarily do so on any particular occasion.

Studies with older subjects, such as Swingley and Aslin (2000) with 18-23 month olds, report that their participants discriminated between minimally different words. Labeled the 'mispronunciation effect,' the experimenters presented one familiar and one a slightly altered ('mispronounced') version of the familiar word e.g., *apple/opple; baby/vaby; ball/gall; car/cur; dog/tog; kitty/pity*. In advance of the experiment, Swingley and Aslin considered that not only comprehension ability but also production ability might affect children's judgements of the test tokens.

> Even if vocabulary size did not provide special leverage in predicting performance, aspects of children's pronuncations of the tested words might be expected to be related to their differentation of good and bad pronuncations in perception. (Swingley and Aslin 2000: 157)

In order to determine to what degree, if any, children's production affected their performance, Swingley & Aslin looked at three different production features relative to the test words: 1) whether children said the words at all; 2) whether they said the words correctly; and 3) whether they said the onset correctly. Interestingly, although considerable evidence for acquisition of more abstract forms is based on children's productions, Swingley and Aslin found that:

> ...the results provided no support for the notions that spoken vocabulary size, or practice in saying the tested words, were related to the size of the mispronuncation effect. (Swingley and Aslin 2000: 158)

Results such as the above from the Swingley and Aslin study strongly suggest that children at slightly older ages are sensitive to featural contrasts, not simply

[19] Infants perform less well on discrimination tasks that involve voiceless fricative pairs like [f] and [θ] (Eilers, Wilson & Moore 1977).

to acoustic-level contrasts (though they are also sensitive to these, of course). As stated earlier, I assume that phonological representations for stored lexical items are made up of feature bundles and that, crucially, individual acoustic patterns for every instance of a lexical item are *not* stored.[20] In the Swingley and Aslin (2000) study, the task presented to the children was one that involved recognition tasks – a match or mismatch of test word (in a sentence) to pictured object. Unlike a discrimination task, this design did not contrast test stimuli immediately with one another, thus the likelihood that the children were judging the tokens on purely acoustic grounds is very low.

The experimental findings cited in this section indicate that infants up to 8 months as well as slightly older children (18-23 mos.) are sensitive to contrasts which correspond to differences in featural representations. Infants in between these ages (i.e., 8-18 mos.) remain sensitive to such contrasts if those contrasts are present in the environment language *and if they are tested using a discrimination task* (Stager and Werker, 1997, e.g., for 14 month olds). On the face of it, this early discrimination ability combined with later discrimination ability would lead one to suppose that initial featural representations were fully-specified – the intervening time is relatively short and all the necessary components appear to be present. However, a number of studies with infants from this intervening age range have found that infants fail to perform well on minimal pair differentiation tasks. These findings are somewhat surprising and so I examine one such study, Werker, Fennell, Corcoran & Stager (2002), in some detail below.

4.4 Against Fully-specified Initial Representations?: Werker, Fennell, Corcoran & Stager, 2002.

4.4.1 Background

Werker, Fennell, Corcoran & Stager (henceforth, WFCS) cite the prevailing view of initial lexical representations in the introduction to the paper. They state:

> Almost all work in child phonology suggests that only gradually, as they acquire words, do children come to represent the more detailed information that might distinguish one word from another [9 studies cited – MJK]. Indeed, during the toddler years, discrimination of minimally different words is only evidence for words that the toddler knows well (Barton, 1980), and even then, perceptual confusions do

[20] For an alternative view, see, for example, Pierrehumbert (2001).

> occur (Eilers and Oller, 1976). All of these studies support the notion that, at the earliest stages of learning a new word, children represent the words only *globally* [emph. MJK]. (WFCS: 2)

While it is not clear what is meant by *globally*, the term is frequently used to refer to children's initial representations. For the immediate purposes, it is only important to note that 'globally' is somehow in contrast with detailed featural representations. WFCS note that, based on past studies, infants in the 11-15 month range seem to be prone to failing on minimal pair differentiation tasks whereas both younger and older children have a much better success rate. They hypothesize that this disparity in performance may be due to some qualitative difference (in, for example, featural representations) or, perhaps, to some task-related effects (where the demands of the task obscured rather than revealed the children's knowledge). The stated aim of the WFCS study is to determine which of these may be true about infants in this age bracket. I provide a summary of the study and its results below.

4.5 Synopsis of the experiment and results:

The participants in the WFCS study were English infants from three different age groups, 14, 17, and 20 months. The test involved an habituation phase during which "...the infant was shown two word-object pairs (e.g., Pair A: word *dih* and object *molecule*, Pair B: word *bih* and object *crown*." (WFCS: 10). The test trials which followed were 'same' or 'switch' trials where, in the former, the infants were presented with the same word-object pairing as in the habituation phase while in the latter, the word-object pairing was switched. WFCS expected that infants who had learned the original pairings (from the habituation phase) would look longer at test trials involving a 'switch' than at those which showed the original pairing (the 'same' trials). The procedures were identical for infants in all three age groups. WFCS found that infants in the 17 and 20 month old groups averaged longer looking times at the 'switch' trials than at the 'same' trials, while the 14 month old infants showed no appreciable difference in looking times at the two different types of trials. The performance of the 14 month old group was not due to an inability to discriminate between the two test tokens *bih* and *dih*, since the authors note that:

> A series of control studies confirmed that infants of 14 months are capable of discriminating these two nonce words ['bih' and 'dih'] in

> a discrimination task that does not allow them to link the words with a nameable object."(WFCS: 3-4)

Minimally, then, the explanation for the performance of the 14 month olds will not be due to failure to perceive distinctions at the acoustic level.

4.6 Discussion

The issue under examination here is whether or not WFCS 2002 provides support for the impoverished initial lexical representation hypothesis. This turns out to be difficult to determine on the basis of the authors' own statements. At one point the authors' attribute the poorer performance of the 14 month olds to *attentional resources* rather than to lack of perception (=featural perception).

> ...the difficulty is explained...on the basis of the attentional resources available to a younger vs. an older infant...
>
> This type of argument is consistent with work in other domains in which it has been shown that the details infants attend to when learning new relations vary depending upon overall task demands (see Cohen, 1998; Cohen & Cashon, 2001; Subrahmanyam, Landau, & Gelman, 1999). According to this argument, the difficulty that the novice word learner faces is not one of having to develop new learning mechanisms or of having to construct a new, lexical based representation. Instead, the difficulty the novice word learner faces is one of being able to detect and encode the detail that is perceptually available at the same time that he/she is attempting to link a newly heard word with an unnamed object. (WFCS: 25)

Attention is a performance factor. The *form* of lexical representations is not affected by attentional demands but, as with other performance effects, competence can easily be masked by performance factors. The implication of the above quote is that the 14 month old *could* store fine phonetic detail if the comptutational demands of the moment were lessened. Words that the child *actually knows* may be richly represented (but those were not the words tested).

The authors also state the following.

> We thus propose that the underlying representation used in both prelexical speech perception tasks and in lexical learning tasks is iden-

> tical, but the infant's ability to utilize all of the information in that representation is compromised through some other limiting fact. (WFCS: 24)

This is a slightly different proposal. This proposal refers explicitly to UR's and implies that the fully-specified UR's are there for the 14 month old (having been established during the habituation phase, presumably) but that they are not accessible due to limitations on computational resources (during the 'same/switch' trials). However, it is not entirely clear that what WFCS intend by 'underlying representation' is what linguists intend by this term. Underlying representations are typically properties only of lexical items – they are not proposed for any 'prelexical' stages of development. From additional discussion about the task facing the infants, it seems clear that WFCS meant something other than 'underlying representations' in a linguistic sense. This is discussed in detail in the following section.

4.7 Association Tasks vs. Knowing Words

The WFCS (2002) study built upon various properties of several earlier studies in terms of experimental design, overall goal of the experiment, and range of ages tested, for example, Werker, Cohen, Lloyd, Casasola, and Stager (1998), henceforth WCLCS, and Stager and Werker (1997) among others. Here and in the earlier studies, the authors discuss the distinction between 'knowing' a word and some other association/recognition phenomena. In WCLCS (1998), the authors state:

> The apparent primacy of comprehension over production raises questions as to whether the cognitive requirements for comprehension are somehow more minimal than those for production. One difference is that comprehension is possible with only recognition memory (hearing a word and recognizing it), whereas production requires recall memory as well (having to retrieve the appropriate word from memory; Huttenlocher, 1974). A more fundamental difference is that some forms of word recognition may reflect the learning of a mere associative link between a word and an object – a "goes together" understanding...(WCLCS 1998:1290)

The present (WFCS) study is not an experiment designed to test representations of words infants 'know.' (And it is certainly not a test of nonce forms that are similar to words that the infants might at some point come to know, given English phonotactics, since *bih* and *dih* are the orthographic equivalents of [bɪ] and [dɪ].) The subjects do not 'know' the test words. 'Knowing' a word is having a representation for it which consists of a stored sound-meaning correspondence. This is true independent of the amount of phonological specification in the stored representation. The subjects in this study (according to their parents) *do* 'know' some words – /bɪ/ and /dɪ/ are not among them. While I see no particular evidence that comprehension has fewer cognitive requirements than production (while it is, on the other hand, fairly obvious that production requires considerably more motor skills than comprehension), it seems logical to assume that the overall memory demands may be different, at least. The task of 'associating' strings of sounds with referents, while a necessary prerequisite to learning a word (i.e., storing a sound-meaning pair in the lexicon), is not, by itself, learning a word. It is perfectly possible, one assumes, to associate some sound that is not a possible lexical item with some referent – the sound of a growl with a dog, for example. WFCS acknowledge this distinction early in the text of their paper:

> ...although associating words with objects is not necessarily equivalent to full referential understanding, we refer to this associative task when used with infants 14 months and older as a word learning task. (WFCS: 4)

However, throughout the paper, starting with the title, 'word learning' appears to the be topic under examination. If the question is really 'Have the 14 month olds developed sound-referent pairings?', the answer would appear to be 'No.' if one takes their performance on the same/switch trials relative to the performance of the older infants to be indicative in this matter. On the other hand, there is no available information about the representations of the words that the 14 month olds are actually supposed to know. A review of the CDI reports might reveal that the infants have minimal pairs in their relatively small comprehension inventories (e.g., 'ball' and 'doll' in the relevant dialects).[21] If that was the case, it might suggest that the infants stored detailed representations (although given parental reporting, that alone would not constitute strong proof). The authors themselves admit this general point readily.

[21] The authors elicited information from the parents regarding the comprehension and production vocabularies of the 14 and 17 month olds using the MacArthur Communicative Development Inventory (CDI).

> Admittedly, it still remains an open question as to whether infants at 14-months would pass or fail to distinguish well known words from minimal pair foils in an on-line recognition task such as that used by Swingley & Aslin (2000), or whether there might be other potentially more sensitive ways to measure the phonetic detail of newly learned words than can be revealed by the "Switch" procedure. (WFCS: 23)

The interesting findings of several of the earlier studies suggest that 14 month olds' performance on association tasks is improved if the nonce forms are not similar to one another phonetically. For example, WCLCS (1998), in a series of experiments similar (though not identical) to the present one, tested 14 month olds on the two nonce forms 'lif' and 'neem'. The infants performed well on the association task using these forms, unlike in the case of the phonetically similar forms 'bih' and 'dih.' The authors of the present study give an account which would explain this disparity.

> Because it is computationally demanding, the infant...has fewer attentional resources to devote to listening carefully to the sound shape of words. Were he/she able to devote more resources to listening, we would argue that he/she would be able to pick up the finest level of detail as is evident in perceptual tasks in infants of the same age. However, because the task involves linking two arbitrary representations...the computational resources are not available and something has to give. In our work, it is the phonetic detail in the words that the infant has trouble attending to closely. (WFCS: 24-5)

This suggests that the difference in performance on phonetically similar and phonetically dissimilar words is in fine vs. gross detail. I take this to be parallel in some sense to visual tasks where, for example, people can report a difference in background color on pictures flashed briefly in front of them (black vs. red) but fail to report accurately whether the small figures on the cards are goats or sheep.

Although the topic under discussion is still performance on 'association tasks' and not representations of known words, this is a rather different conclusion than one which suggests that infants of this age may be able to store phonetic detail but simply not access that detail on a particular task. This type of attentional explanation actually prevents infants from *storing* detail. In an actual test of 'known' words, this would be equivalent to concluding that the infant does not store fully-specified UR's at the age of 14 months. Note that the ability to perceive (at the acoustic level) fine phonetic detail is only one of three critical steps in the process. The infant must, in addition, be able to *attend* to those details and further, must be able to *store* those details in a featural representation form. Can one take it, then, that this particular explanation of the performance on association tasks is indicative of a lack of detail in the infants' 'known' word representations? I argue in the following section that the answer to this question is also 'No.'

4.8 The Task and the Individual

Leaving aside the differences between association tasks and knowing words, two points which I feel are crucial in evaluating the experimental results require further discussion. The first of these is whether the response of the infants is primarily due to task-related factors or not.

As noted in the synopsis of the experiment, WFCS hypothesize that poor performance of 14 month olds on earlier, similar experiments may have been due to task-related factors. Thus, in the current study, WFCS use more physically disparate objects and allow increased exposure time with the goal of making the task less difficult. Their concluding remarks recap their view about the lack of impact these changes had on the performance of the 14 month olds.

> ...the difficulty is explained not on the basis of amount of exposure, but instead on the basis of attentional resources available to a younger vs. an older infant even in the face of equivalent amounts of exposure. (WFCS: 25)

For the authors, task-related factors are apparently limited to the parameters of a *particular* task. However, one cannot conclude that performance is not still due primarily to the task just based on this. Changing the particulars of a task is not equivalent to changing the task itself. The task, as far as I can determine, was exactly the same as in the earlier experiments, i.e., it was an habituation period followed by 'same/switch' trials. An actual task change would be going from a discrimination task to a recognition task, for example. Without a change of task, it is difficult to see how one can rule out the task as the impediment to better performance. Moreover, given the dramatic developmental changes that take place during the ages in question (between the first and second birthdays, broadly), it seems likely that a task which might be trivial for a 17 or 20 month old would still be impossible for a 14 month old. To highlight disparities in developmental stages, 'amount of exposure' may well be a case in point. If the 14 month olds actually do know some words (as reported), then apparently, given sufficient exposure (of some type), they were able to store some sound-meaning correspondences and already, therefore, have a lexicon. While it is logically possible that they developed this lexicon purely based on some different 'type' of exposure and that 'amount' of exposure was not a factor, this seems unlikely, on the surface. Differences in developmental periods must, however, certainly be the relevant factor in the difference in performances *within* the 14 month old group itself.

As mentioned, WFCS also measure vocabulary size of the infants in the study. It is in the discussion of the correlation between vocabulary and performance on the test that WFCS state:

> Those infants with productive vocabularies of $<$ 25 words looked equally to same and to switch trials, whereas those infants with productive vocabularies of more than 25 words looked significantly longer to the switch trials,.... The latter group includes all of the infants 20 months of age, nine of the infants 17 months, and four of the infants at 14 months of age. (WFCS: 19)

I point to this particular prose statement to emphasize the fact that the authors make it perfectly clear, not only by citing the numerical results but also in prose, that the performance of the 17 month olds and 14 month olds was not uniform across infants. Some of the 17 month olds performed as poorly as the 14 month olds and, perhaps more tellingly, some of the 14 month olds performed as well as the older infants on the task. This is expected, given that actual age is not perfectly correlated with development. The overall results, as they are discussed in the 'General Discussion' section and in the 'Summary,' are interpreted based on the performance of the *group* – a normal feature of psychological experiments. The difficulty, as I see it, lies in the fact that linguistic knowledge is a matter of the individual.

4.9 Interpreting the Results

I return to the question of whether this study is support for impoverished or, alternatively, fully-specified initial featural representations. The authors themselves do not take a position *explicitly* on this matter. They do indicate that they are taking a position contra Rice and Avery (1995) but describe such a position as follows:

> The strongest 'discontinuity' models are those in which it is argued that the representation used in lexical (word-learning) tasks is distinct from the representation used in perceptual tasks (e.g. Brown &

> Matthews, 1997; Keating, 1984; 1988; Rice & Avery, 1995; Shvachkin, 1948/1973). All these models share the feature of requiring new phonological representations, distinct from those used in perception, to be assembled after the child begins to learn meaningful words. They argue that these two distinct representations continue even after word learning is complete and thus predict less phonetic detail in word recognition than in speech perception tasks. (WFCS: 24)

Specifically, the authors seem to be focussing primarily upon the issue of whether the perceptual information continues to be available to the infant such that it can ultimately be used in lexical representations. The point at which the infant is able to use this detailed information is the point when the infant's computational resources are sufficient to allow simultaneous processing of the phonetic detail and other demands of word learning. This appears to be what the authors' intend by a 'developmental' explanation – the explanation that they favor. WFCS have been careful throughout their paper to consider alternative explanations and such a conclusion is, I believe, in keeping with this, since it leaves the question of initial representations open. As I stated in the Introduction, one of my reasons for choosing the WFCS study was that it could be seen to be having an impact. It is, in fact, the most recent of a line of studies and simply serves to confirm findings from previous studies regarding the same phonetically similar pairs. However, other researchers are citing the WFCS study as evidence to support impoverished initial representations. In fact, Pater (to appear), a co-author from a previous, related study states of Stager and Werker (1997):

> A reasonable interpretation of the infants' failure to respond to the switch in place of articulation in Experiment 1 is that at this stage of development, the place distinction is not encoded in lexical representations, so that words differing only in this respect are treated as identical in tasks involving meaning. (Pater, to appear: 7)

Similarly, Fikkert and Levelt (2002:1-2) state of Stager and Werker (2000):

> We challenge the common assumption that the child's representations are essentially adult-like, a claim which is largely based on instances where it is clear that the child perceives more than he is able to produce (such as the 'fish'-phenomenon), and on early infant perception studies. However, recent studies (cf. Werker & Stager, 2000) have shown that as soon as *linguistic* perception (referring to stored lexical representations) is taken into account, the story may be different. . .

The above interpretations of this line of study do not have a foundation in the studies themselves, in my opinion, for the many reasons I have cited above. One cannot, of course, blame the authors for the failure of other researchers to read the study or consider its results carefully.

5 Conclusion

So do experimental studies inform theoretical research? The answer seems to be that while there is interaction between the two domains, that interaction does not appear to be as useful and productive as it could be. Both of these groups share a common goal – to understand the nature of language acquisition. Currently, however, a situation seems to exist in which 'a little knowledge is a dangerous thing.' As this paper illustrates, the amount of truly shared knowledge between the two groups appears to be minimal. Experimental researchers use some of the same terminology as theoretical researchers, but not with the same meaning. Correspondingly, one feels that there is, on occasion, a lack of clarity about what exactly is being tested and why, and how any results can contribute to the bigger picture of our understanding of language. Meanwhile, theoretical researchers are apt to ignore all the experimentalists' cautious statements about their findings and are, instead, quick to exploit superficial details in pursuit of their own research agendas. Lack of training in experimental design and the interpretation of results as well as lack of knowledge of the general framework within which experimental researchers work makes a genuine evaluation of experimental studies both difficult and time-consuming. This, however, does not excuse theoretical researchers from making a thorough evaluation of experimental studies and sharing some of the caution of the authors of such studies in interpreting the results. It is clear that when a superficial, rather than a thorough, reading of experimental results is transmitted across the intellectual landscape as anecdotal information (i.e., without the experimental details), the snowball effect can be considerable.

References

Barton, D. 1980. 'Phonemic Perception in Children,' in G.H. Yeni-Komshian, J. Kavanaugh, and C.A. Ferguson (eds.) *Child Phonology: Volume 2 - Perception*. New York: Academic Press, 97-116.

Barton, D. 1978. 'The Discrimination of Minimally-Different Pairs of Real Words by Children Aged 2;3 to 2;11.' in N. Waterson & C. Snow (eds.) *The Development of Communication*. Chichester: Wiley, 255-261.

Best, C.T., G.W. McRoberts, R. LaFleur & J. Silver-Isenstadt. 1995. 'Divergent Developmental Patterns for Infants' Perception of Two Nonnative Consonant Contrasts.' *Infant Behavior and Development*, 18: 339-350.

Best, C.T., G.W. McRoberts & N.M. Sithole. 1988. 'Examination of the Perceptual Re-Organization for Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants.' *Journal of Experimental Psychology: Human Perception and Performance* 13: 345-360.

Brown, C. & J. Matthews. 1991. 'The Role of Feature Geometry in the Development of Phonemic Contrasts,' in S.J. Hannahs and M. Young-Scholten (eds.), *Focus on Phonological Acquisition*. Amsterdam: John Benjamins. 67-112.

Burnham, D.K. 1986. 'Developmental Loss of Speech Perception: Exposure to and Experience with a First Language,' *Applied Linguistics*, 7: 201-240.

Cohen, L.B. 1998. 'An Information-Processing Approach to Infant Perception and Cognition.' in F. Simion & G. Butterworth (eds.) *The Development of Sensory, Motor and Cognitive Capacities in Early Infancy: From Perception to Cognition* Hove, UK: Psychology Press/Erlbaum, 277-300

Cohen, L.B. & C.H. Cashon. 2001. 'Infant Object Segregation Implies Information Integration.' *Journal of Experimental Child Psychology* 78: 75-83.

Eilers, R.E. & D.K. Oller. 1976. 'The Role of Speech Discriminations in Developmental Sounds Substitutions,' *Journal of Child Language* 3: 319-329.

Eilers, R.E., W.R. Wilson, & J.M. Moore. 1977. 'Developmental Changes in Speech Discrimination in Infants.' *Journal of Speech and Hearing Research*, 20: 766-780.

Eimas, P.D., E.R. Siqueland, P.W. Jusczyk & J. Vigorito. 1971. 'Speech Perception in Infants.' *Science* 171: 303-306.

Fikkert, P. & C. Levelt. 2002. 'Putting Place into Place,' abstract for the 25th GLOW Colloquium, Amsterdam.

Hale, M. & M. Kissock. 1997. 'Nonphonological Triggers for Renewed Access to Phonetic Perception,' in A. Sorace, C. Heycock, and R. Shillcock (eds.) *Proceedings of the GALA '97 Conference on Language Acquisition*,' University of Edinburgh, 229-234.

Hale, M., M. Kissock & C. Reiss. ‘Microvariation, Variation, and the Features of Universal Grammar,’ *Language and Speech* (to appear). [A version of this paper was presented at the 25th GLOW Colloquium, April 2002, Amsterdam.]

Hale, M. & C. Reiss. 2002. ‘The Subset Principle in Phonology: Why the *Tabula* Can’t Be *Rasa*’. ms.

Harnad, S. (ed.) 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Huttenlocher, J. 1974. ‘The Origins of Language Comprehension,’ in R.L. Solso (ed.), *Theories in Cognitive Psychology*. Potomac, MD: Erlbaum. 331-368.

Inkelas, S. 1994. ‘The Consequences of Optimization for Underspecification.’ Rutgers Optimality Archive #40-1294.

Jakobson, R. 1941. *Kindersprache, Aphasie und Allgemeine Lautgesetze.* Selected Writings, vol. 1, 328-401. The Hague: Mouton, 1971.

Jusczyk, P.W. & R.N. Aslin. 1995. ‘Infants’ Detection of the Sound Patterns of Words in Fluent Speech.’ *Cognitive Psychology* 29: 1-23.

Jusczyk, P.W. & P.A. Luce. ‘Speech Perception and Spoken Word Recognition: Past and Present.’ *Ear and Hearing* in press.

Keating, P.A. 1988. ‘Underspecification in phonetics,’ *Phonology* 5: 275-292.

Keating, P.A. 1984. ‘Phonetic and Phonological Representation of Stop Consonant Voicing,’ *Language* 60: 286-319.

Kingston, J. & R. Diehl. 1994. ‘Phonetic Knowledge,’ *Language* 70: 419-454.

Kiparsky, P. & L. Menn. 1977. ‘On the Acquisition of Phonlogy,’ in J. Macnamara (ed.) *Language Learning and Thought*. New York: Academic Press, 47-78.

Kissock, M. & M. Hale. 1998. ‘Nonphonological Triggers for Renewed Access to ‘Phonetic’ Perception,’ in *Language Acquisition: Knowledge Representation and Processing*. Proceedings of GALA ’97, University of Edinburgh, 1998.

Kuhl, P.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, P.K. 1979. ‘Speech Perception in Early Infancy: Perceptual Constancy for Spectrally Dissimilar Vowels.’ *Journal of the Acoustical Society of America*, 66:1168-79.

Liljencrants, J. & B. Lindblom. 1972. ‘Numerical Simulations of Vowel Quality Systems: The Role of Perceptual Contrast,’ *Language,* 48: 839-862.

Lisker, L. & A. Abramson. 1970. ‘The Voicing Dimension: Some Experiments in Comparative Phonetics,’ *Proceedings of the 6th International Congress of PHonetic Sciences,’* Prague: Academia. 563-567.

Maye, J. & L. Gerken. 2000. ‘Learning Phonemes Without Minimal Pairs.’ in *Proceedings of the 24th Annual Boston University Conference on Language Development*.

Maye, J., J.F. Werker & L. Gerken. 2002. ‘Infant Sensitivity to Distributional Information Can Affect Phonetic Discrimination.’ *Cognition* 82: B101-B111.

Menn, L. & E. Matthei. 1992. ‘The “Two-Lexicon” Approach to Child Phonology: Looking Back, Looking Ahead,’ in C.A. Ferguson, L. Menn, and C. Stoel-Gammon (eds.) *Phonological Development: Models, Research, Implications.* Timonium, MD: York Press.

Pater, J. ‘The Comprehension/Production Dilemma and the Development of Receptive Competence,’ in R. Kager, J. Pater, and W. Zonneveld (eds.) *Fixing Positions: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press. (To appear)

Pierrehumbert, J. 2001. ‘Stochastic Phonology, *GLOT International* 5.6: 195-207.

Polka, L. and J. F. Werker. 1994. ‘Developmental Changes in Perception of Non-Native Vowel Contrasts.’ *Journal of Experimental Psychology: Human Perception and Performance* 20:421-435.

Rice, K. & P. Avery. 1995. ‘Variability in a Deterministic Model of Language Acquisition.’ in J. Archibald (ed.) 1995a. *Phonological Acquisition and Phonological Theory.* Hillsdale, NJ: Lawrence Erlbaum Associates. (pp. 23-42)

Shvachkin, N.K. 1973. ‘The Development of Phonemic Speech Perception in Early Childhood,’ in C. Ferguson and D. Slobin (eds.) *Studies of Child Language Development*. New York: Holt, Rinehart & Winston. 91-127. (Original work published 1948)

Smolensky, P. 1996. 'On the Comprehension/Production Dilemma in Child Language,' *Linguistic Inquiry* 27:720-731.

Stager, C.L. & J.F. Werker. 1997. 'Infants Listen for More Phonetic Detail in Speech Perception Than in Word Learning Tasks,' *Nature*, 388: 381-382.

Stevens, K.N. 1972. 'The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data,' in P.Denes and E. David, Jr. (eds.) *Human Communication: A Unified View*. New York: McGraw Hill.

Subrahmanyam, K., B. Landau & R. Gelman. 1999. 'Shape, Material, and Syntax: Interacting Forces in Children's Learning in Novel Words for Objects and Substances.' *Language and Cognitive Processes* 14: 249-281.

Swingley, D. & R.N. Aslin. 2000. 'Spoken Word Recognition and Lexical Representation in Very Young Children,' *Cognition* 76:147-166.

Trehub, S.E. 1976. 'The Discrimination of Foreign Speech Contrasts by Infants and Adults.' *Child Development* 47: 466-472.

Tsushima, T., O. Takizawa, M. Sasaki, S. Siraki, K. Nishi, M. Kohno, P. Menyuk, & C.T. Best. 1994. 'Discrimination of English /r-l/ and /w-y/ by Japanese Infants at 6-12 Months: Language Specific Developmental Changes in Speech Perception Abilities'. Paper presented at the International Conference on Spoken Language Processing 4. Yokohama, Japan.

Werker, J.F. 1994. 'Cross-Language Speech Perception: Developmental Change Does Not Involve Loss,' in J. Goodman and H. Nusbaum *The Development of Speech Perception*. Cambridge: MIT Press. 93-120.

Werker, J.F., L.B. Cohen, V.L. Lloyd, M. Casasola & C.L. Stager. 1998. 'Acquisition of word-object Associations by 14-month-old Infants.' *Developmental Psychology* 34: 1289-1309.

Werker, J.F. & J. S. Logan. 1985. 'Cross-Language Evidence for Three Factors in Speech Perception,' *Perception and Psychophysics* 37: 35-44.

Werker, J.F., C.T. Fennell, K.M. Corcoran & C.L. Stager. 2002. 'Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size.' *Infancy* 3: 1-30.

Werker, J.F., J.H. Gilbert, K. Humphrey & R.C.Tees. 1981. 'Developmental Aspects of Cross-Language Speech Perception.' *Child Development* 52: 349-355.

Werker, J.F. & C.L. Stager. 2000. 'Development Changes in Infant Speech Perception and Early Word Learning: Is There A Link?,' in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology 5* Cambridge: Cambridge University Press. 181-193.

Werker, J.F. and R.C. Tees. 1984. 'Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life,' *Infant Behavior and Development* 7: 49-63.

Yip, M. 1996. 'Lexicon Optimization in Languages Without Alternations.' Rutgers Optimality Archive #135-0996.

Madelyn Kissock
Dept. of Classics, Modern Languages, and Linguistics
1455 de Maisonneuve West, H 663-15
Concordia University
Montréal, Quebec, H3G 1M8
Canada
mkissock@alcor.concordia.edu