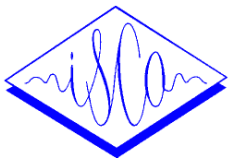

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

RELATIONSHIP BETWEEN SPEECH PERCEPTION AND PRODUCTION IN LANGUAGE ACQUISITION

Paula Menyuk

Boston University
Boston, MA 02215, U. S. A.

ABSTRACT

The relation between speech perception and production during the period of lexical acquisition is still a matter of much discussion but little resolution. There have been three principal models that have been proposed to explain possible relations between perception and production. These models, the data they account for and the questions they leave unanswered will be discussed. A fourth model will be presented which attempts to explain the relation between speech perception and production when taking into account other aspects of language, and all possible constraints on the process of mapping perceptions into productions. The model will be tested by examining how well it can account for different types of findings about young children's speech production.

I. INTRODUCTION

The known facts about speech perception and production during the pre-lexical period indicate that there are different physiological and psychological mechanisms that account for these behaviors. Speech production seems to be under the governance of the developing control of the vocal system, while speech perception seems to be under the control of the auditory system, and the development of memory. The speech behavior of the pre-lexical infant is driven by those systems that are innately given to the human infant; that is, the ability to produce articulated sound, and to perceive and remember acoustic differences that are categorical in nature. Thus speech perception and production do not appear to be related at all during this period. However, at about ten months of age, the infant's speech behavior, apparently, begins to be based on the particular language of the environment. Infants begin to be most sensitive to the speech sound differences of their language, and the prosody of their babbling begins to resemble the patterns of native speakers in their environment [1]. It is at this point in development that speech perception must be playing a guiding role in speech production. These observations, however, do not reveal to us what the relation is between speech perception and production, but only that such a relation must exist.

II. MODELS

There have three principal models that have been presented to account for speech behavior in the lexical period. One model suggests that speech production is a subset of what the child perceives [2]. The child perceives many if not all of the speech sound distinctions that she has difficulty producing. Thus, for example, the child might perceive the difference between the words "pat" and "bat" but have difficulty in producing them distinctly. The subset that is produced can be accounted for by the mapping of acoustic information into motor movements via a set of realization rules that the child develops. There is ample evidence to support this notion since children frequently appear to understand differences they do not productively realize but it is not clear that the child is perceiving speech sound distinctions the way an adult does.

Another model suggests that all that is perceived is mapped but that the perception is distorted in some way, and that this distortion is reflected in the production. Thus, the distinction between "pat" and "bat" might not be the same for the child and the adult, although the child does perceive some distinction. It is very difficult to test this model, but there is some evidence that children may distinguish between speech sounds on the basis of acoustic parameters that differ somewhat from those used by adults [3]. It is not clear, however, that children's unique productions are always based on unique perceptions of these words. It has been hypothesized by a number of researchers that speech production representations may be distinct from those of speech perception [4].

A third model suggests that both speech perception and production are governed by cognitive or processing constraints [5]. Each system, perception and production, is constrained by the nature of the particular system involved in the perception or production act. For example, perceiving the distinction between "pat" and "bat' might not be difficult in terms of acoustic feature distinctions between the two sounds, but difficult to produce because of the timing controls on the vocal mechanism that are required for that distinction. Other sound distinctions might be both difficult to perceive and produce. For example, the distinction between "bite" and "bright" might be both difficult to perceive and to produce at some stage of development; difficult to perceive because of the amount of information in the opening segments, but difficult to produce because of the rapid motor movement requirements of consonant clusters.Other distinctions might be easy to both perceive and produce. "Mama" versus "Papa" may be such a distinction. This model might account for some of the apparent randomness that

has been observed in the early speech production of children. While the last suggestion seems to be a reasonable one given what is known about each of the systems, and what has been noted about avoidance and idiosyncratic speech production behavior, it is unsatisfactory in that it ignores what must be the case. For the most part, children must rely on their speech perception to produce lexical items that are either recognizeable or consistent enough to be identified.

Each of the models proposed above seems to be partially explanatory in that each model accounts for some aspects of the speech behaviors of children during the lexical acquisition period. Perhaps, then, a model that combines each might be most explanatory. However that seems to be a less than satisfactory solution for two reasons. First, what all the models have not accounted for in a satisfactory manner is that speech perception and production during the lexical period must be governed by higher order nodes, the word or even the phrase. Lexical meaning and frequency of lexical items must play a role in how the child maps perceptions into productions. Further, most children must approach the task of mapping by more than one translation strategy, and have, as their end goal, different products on different occasions: the prosodic pattern, the word or the phrase. Thus, early utterances may be composed of sequences of unintelligible sounds, that are marked prosodically, or recognizable words or phrases. The recognizable words and phrases are also marked with an appropriate prosodic pattern to indicate the intent of the utterance. There are, then, additional sources that control the mapping of speech from perception to production that are independent of the speech perception and production mechanisms themselves. Some of these additional sources are word meaning, word frequency, the target of the production (intonational contour or intonational contour plus word or plus phrase) and the intent of the utterance. Thus, the child does not just map an acoustic signal into a sequence of motor movements. The child is attempting to map messages perceived into messages that will convey intent.

The second problem with assuming that all of the models that have been proposed do, in fact, account for aspects of the speech behavior of children during the lexical acquisition period is that the assumption might make it very difficult to account for universals of phonological development. There have been many studies of speech perception and many inventories of the speech production of children during the early stages of language acquisition that suggest a universal sequence of acquisition of perceptual and productive speech sound distinctions. However, if the bases for mapping speech perceptions to speech productions are as plentiful as suggested, it seems highly unlikely that such a universal sequence actually exists. What may be the case, instead, is that the many bases for the mapping process are ordered, and that, therefore, there is a seeming order in the products of the mapping process both perceptually and productively. To determine whether or not this is the case, the data that have been collected on children's speech perception and production during the lexical acquisition period will be very briefly reexamined.

III. TESTING MODELS

Given all that is known about infants one might suggest that the initial approach used by the child in mapping perception to production is to pay attention to semantically important lexical items or phrases that are used by others in the environment. This would dictate the domain for application of additional approaches. If we examine the early vocabularies of most children we find that certain semantic categories are always represented. The categories include words for important caregivers (perhaps "momma" and "poppa"), foods (names for water, milk, cookie, etc., actions ("up", "down", and affect ("no", "nice", "good", dirty, etc.). Thus both meaning and frequency are calling the child's attention to specific lexical items. In addition, these lexical items may be said with emphasis, which would make them perceptually clear in comparison with other words. Among the important, frequently heard and possibly emphasized words are lexical items that are easy to map into production realizations because they require simple closing and opening of the vocal tract, movements ofen engaged in during the babbling period. The words "no", "momma" and "papa" fall into this category, and are early and well produced lexical items. What follows after this are possible approaches to the realization of the other lexical items in the child's repertoire.

There are several general "rules" of mapping from perception to production that have been pointed to by a number of researchers. These rules include the following: vowel harmony, syllable repetition, weak syllable deletion, backward and forward assimilation, avoidance, consonant cluster reduction, idiosyncratic realizations, stopping continuants, preferences, etc. All of the above "rules" imply that the child is trying to make a good approximation to the lexical items he or she perceives in part or in whole. This will be discussed further below, but for the moment we will simply examine the application of these mapping rules.

First, all the rules, obviously, do not apply to every item in the child's repertoire. They can only apply to lexical items on which the rules can operate.For example, rules concerning mutisyllable treatment can only apply to multisyllable words. Therefore, these rules will come later in development except for semantically important (to the child) items (for example, the words "cookie" and "water" in the examples given above). Under these circumstances some of the syllabic rules will apply and the child will produce approximations (such as /tutu/, and /wawa). The point is that rules will only apply to selected lexical items, thus the first step is selection. The second step after selection is approximation using the easiest translations possible. The rules applied in the approximation examples above are fronting (/t/ for /k/ in "cookie" and /w/ for /t/ in "water") and syllable repetition. Each of the lexical items in the repertoire will be approximated with somewhat

different rules depending on the shape of the lexical item that is the target, but the general rules seems to be- approximate using the easiest motor movements possible without totally losing the target. The nature of the approximation, as in the examples given above, will indicate the order in which the rules apply for a given child. The important point is that message choice, and lexical choice within the message governs speech perception and production.

After making lexical choices, the application of rules in the above examples could imply that the child is perceiving the lexical item in the way the adult does but simply cannot produce it the way the adult does (model 1 applies). However, each of the other two models could account for some of the data that points to "rule" application as well. For example, the rule of consonant fronting might be due to the child not perceiving the appropriate formant transitions from the /k/ to the vowel in "cookie" the way the adult does but perceiving the stop element in the consonant. Therefore, the child fronts with a stop not a nasal (model 2 seems to apply in this instance). Or it is possible that the child perceives some aspects of the back consonant transition into the vowel but only can translate this into particular timing and aspiration cues and produces the more front consonant when attempting to execute the multisyllable word (model 3 applies). All of these possibilities are speculations. The realization of the word tells us, indeed, that the stop element of the consonant is preserved as is the vocalic element, but neither of these segments is accurately reproduced. Thus there is little evidence to support any one of the three models in these early word productions.

To lend support to the notion that syllabic structure is preserved in these early utterances, one finds that in phrase as well as lexical production, there is preservation of syllabic structure and categorical features of segments as well. For example, such early phrases as "wanna chu" (want a shoe), and "pud a ded" (put on desk) certainly preserve the syllabic structure of the model utterance, and roughly approximate the segmental structure. In addition, the intonation and stress of the model utterance is preserved. Word recognition at this stage may be largely based on perceiving these gross representations and the contexts in which they are produced, and recoding these perceptions directly into productions.

Somewhat later both productive and perceptual behavior indicate that a transition has occurred. These later behaviors further suggest that the earlier ones are probably not based on accurate perceptual distinctions between syllabic units but, rather gross ones. For example, in early consonant cluster reduction what appears to be preserved are single segments of the cluster. The word "milk" is produced as /muk/. Later, the child's productions more closely approximate the adult form. This might be due to the child's perceiving one of the consonantal segments plus something, but not perceiving both segments the way the adult does (model 2) or perceiving some continuant segment in strident clusters (/s/ + stop as in the word "stop") or a rounding segment in clusters containing glides (stop + /r/ or /l/) as in "blue" and "grow") and mapping these features into motor movements. Model 2 does not fully account for the fact that there appear to be particular perceptions of the plus something. This is indicated by the fact that there are particular realizations of these words now. The word "stop" might be realized as /tsop/ and the word "blue" as either /buu/ or /bwu/ [6].

These data indicate that model 3 is more explanatory at this stage than model 2. There is little evidence to support model 1 at either of these two early stages. If model 3 is explanatory at this stage then there appears to be a closer relation between perception and production than model 3 initially proposed. Further evidence that there is now a closer relation between perception and production comes from observations that children at this later stage will not accept adults' approximations to their own productions

STAGE I
GROSS APPROXIMATIONS

CYCLE 1- PAY ATTENTION TO SEMANTICALLY IMPORTANT LEXICAL ITEMS (USUALLY STRESSED)

CYCLE II- SELECT FOR REPRODUCTION LEXICAL ITEMS THAT ARE EASY TO TRANSLATE

CYCLE III- MAKE A GOOD APPROXIMATION OF CLEAR CV, CVCVs

CYCLE IV- PERCEIVE THE SYLLABIC AND CATEGORICAL FEATURES OF LEXICAL ITEMS ALONE AND IN SEQUENCES

CYCLE V- PRODUCE ITEMS ALONE AND IN PHRASES WITH APPROPRIATE PROSODY AND PRESERVING CATEGORICAL FEATURES

STAGE II
CLOSER APPROXIMATIONS

CYCLE I- PAY ATTENTION TO DETAILS OF SYLLABLE STRUCTURE

CYCLE II-TRANSLATE THESE DETAILS AS ACCURATELY AS YOU CAN

CYCLE III- PAY ATTENTION TO THE DETAILS OF INITIAL SEGMENTS OF SYLLABLES

CYCLE IV- TRANSLATE THESE DETAILS AS ACCURATELY AS YOU CAN

STAGE III
AN APPROXIMATION TO THE ADULT MODEL

CYCLE I- PERCEIVE SPEECH IN CONTEXT THE WAY THE ADULT DOES

CYCLE II- PRODUCE WHAT YOU PERCEIVE ACCORDING TO YOUR SYSTEM OF RULES

Fig. 1- Descriptive model of development of the relation between perception and production of speech at the early stages.

as good examples of a lexical item. The child may say something approximating /fis/ for "fish" but will only accept /fish/ as correct.

None of the models discussed predicts the possibility that the relation between perception and production changes during the period of early lexical acquisition. Above in Fig. 1 a descriptive model of the proposed developments between perception and production of speech during the early lexical period is presented. This model attempts to incorporate factors that may govern the relation between speech perception and production at early stages. The model also attempts to describe and order changing relations between speech perception and production over time.

At stage I of the developmental process, gross approximations to both perceptual representations and productive representations take place. At stage II finer distinctions are made both in the perceptual and productive process. Each stage is governed by the motivation to communicate effectively, and at each stage there is a close relation between perception and production with production always based on perception. The description of the final stage, stage III, attempts to account for the unique rules that children develop by positing that children, at this stage, do perceive speech they attend to in the way the adult does but produce it according to the general and idiosyncratic rules they have developed. Perception and production are no longer so intimately tied.

The problem with positing a close relation between perception and production at this later point is that there are some instances of the child branching off from perceptions to come up with unique realizations for particular lexical items. Examples of this behavior are so-called "non-isomorphic processes" and "preferences" [7]. A possible explanation for these behaviors may be the sudden spurt in the child's lexical development, and the resultant stress that is placed on the storage and retrieval capacities of the young child. At the stage when the words "dress" and "blue" appear in the child's lexicon there are probably five times as many lexical items in the child's repertoire than there were at the earlier stage. From about twelve to eighteen months that lexicon grows from approximately ten words comprehended and produced to about one hundred words comprehended and about fifty words produced. From eighteen months to twenty four months there is a ten-fold increase in the number of lexical items in a normal child's repertoire. Further, the child's semantactic and pragmatic development is, of course, going on simultaneously and at the same hectic pace.

These pressures also might account for the fact that there is not a consistent development of finer and finer approximations to the adult model in terms of production. There is apparent backsliding for some lexical items. Some of the substitutions for segments that were used earlier in certain words and then refined (for example, /dwes/ becomes "dress") are applied in new instances (for example, as "glass is acquired it is pronounced /gwas/). So-called frozen forms remain in the productive lexicon ("doggie" remains /gogi/ for a long time), and, as stated, preferences for sound realizations occur (/nana/ for "banana" and /neni/ for "penny") and idiosyncratic productive realizations occur (for example, /munugeti/ for "spaghetti"). Similar backsliding, frozen forms, etc. occur in other aspects of language at this time. In addition to the pressure of constantly acquiring new words and structures, there is the pressure of acquiring words and structures that are more complicated. In terms of words, there is an increase in the number of multisyllabic words that differ in CV structure across the word, and an increase in words that have more complex consonant clusters initially and finally.

Despite this seeming randomness in mapping of perception to production, the child institutes some regularity into the system by doing the best he or she can under these difficult circumstances. The phonological rules discussed above are applied to bring about some regularity into the system. It may be at this point that perceptual and productive representations may be drawing apart, and Model 1 might best account for the data. There is a great deal of evidence, primarily from brain damaged individuals, that, in the adult, speech perception and speech production representations are stored in separate parts of the brain. Only further research will tell us whether or not this process is beginning in the child at this stage of development.

IV. REFERENCES

[1] B.Boysson-Bardies, L. Sagart, & C. Durand, "Discernible differences in the babbling of infants according to target language." J. Child Language, 11, 1-15, 1984.

[2] N. Smith, "The acquisition of phonology: A case study." Cambridge: Cambridge U. Press,1973.

[3] S. Nittrouer, M. Studdert-Kennedy, R. McGowan, "The emergence of phonetic segments: Evidence from the spectral structure of fricatives in novel syllables spoken by children and adults". J. Speech and Hearing Research, 32, 120-132, 1989.

[4] P. Menyuk, L. Menn, & R. Silber, Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.)," Language Acquisition, second edition", (pp. 198-222). Cambridge: Cambridge U. Press, 1986.

[5] C. Ferguson, C., Discovering sound units and constructing sound systems, In J. Perkell & D. Klatt (Eds.) "Invariance and variability in speech processes", (pp. 36-51). Hillside, N.J.: Erlbaum, 1986.

[6] P. Menyuk, & D. Klatt, " Child's production of initial consonant clusters", Quarterly Progress Reports, Research Laboratory of Electronics, M.I.T., No. 51, 205-213, 1968.

[7] D. Ingram, Phonological development: production. In P. Fletcher & M. Garman (Eds.) "Language acquisiton, second edition," (pp. 223-239). Cambridge: Cambridge U. Press, 1986.