A descriptive study of the motor theory of speech perception

Wendy Maureen Coburn

The University of Montana
A DESCRIPTIVE STUDY OF THE MOTOR THEORY OF
SPEECH PERCEPTION

By

W. Maureen Coburn
L.C.S.T., Edinburgh, Scotland, 1969

Presented in partial fulfillment of the requirements
for the degree of

Master of Arts
UNIVERSITY OF MONTANA
1972

Approved by:

Chairman, Board of Examiners
Dean, Graduate School

Date
ACKNOWLEDGMENTS

The writer wishes to thank her thesis committee for time and advice concerning this investigation. Gratitude is also expressed to the school principals and speech therapists with whom she was in contact, the secretaries in the Speech and Hearing Clinic, and fellow graduate students who were of assistance in the preparation of this thesis. A special thanks is given to Debbie, who drew the illustrations for my test, and to Kitty, who lent me her "ears" and a lot of her time. Also, for moral support and incentive, I thank Frank and Demois.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>STATEMENT OF THE PROBLEM</td>
<td>10</td>
</tr>
<tr>
<td>2. PROCEDURE</td>
<td>14</td>
</tr>
<tr>
<td>SUBJECTS AND SELECTION CRITERIA</td>
<td>14</td>
</tr>
<tr>
<td>MATERIALS USED</td>
<td>17</td>
</tr>
<tr>
<td>TESTING PROCEDURES</td>
<td>18</td>
</tr>
<tr>
<td>ANALYSIS OF DATA</td>
<td>21</td>
</tr>
<tr>
<td>3. RESULTS</td>
<td>22</td>
</tr>
<tr>
<td>4. DISCUSSION</td>
<td>25</td>
</tr>
<tr>
<td>RECOMMENDATIONS</td>
<td>34</td>
</tr>
<tr>
<td>5. SUMMARY AND CONCLUSIONS</td>
<td>38</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>40</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>43</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>50</td>
</tr>
</tbody>
</table>
LIST OF TABLES AND FIGURES

Table

1. P.P.V.T. Scores and Ages of Subjects ........ 16

Figure

1. Scores for All Ten Subjects for Both Tapes and Number Correct for /r/ and /w/ Words for Both Tapes ........ 23
Chapter 1

INTRODUCTION

Lastly, I am to take notice, that there is so great a Communication and co-respondency between the Nerves of the Ear, and those of the Larynx, that whensoever any sound agitates the Brain, there flow immediately spirits towards the Muscles of the Larynx, which duey dispose them to form a sound altogether like that which was just now striking the Brain. And although, I well conceive, that there needs some time to facilitate those motions of the muscles of the Throat, so that the Sounds, which excite the Brain the first time, cannot be easily expressed by the Throat, yet notwithstanding I doe as well conceive, that by virtue of repeating them it will come to pass, that the Brain, which thereby is often shaken in the same places, sends such a plenty that at length they easily move all the cartilages, which serve for that action, as 'tis requisite they should be moved to form Sounds like those, that have shaken the Brain.

It was in 1668 that de Cordemoy presented the above philosophical discourse concerning speech. Even then the interdependence of speech production and perception was recognized and might be considered a forerunner of the current idea that "production follows perception." This idea is in opposition to the one that states that "perception follows production" put forth by the Haskin's group.

Advocates of the motor theory of speech perception, i.e., "perception follows production" (Liberman, Cooper, Harris, and MacNeilage, 1963), contend that "speech sounds are perceived by reference to the articulatory movements
that produce them and this articulatory reference is important for the distinctiveness of speech as perceived."

According to this view, the sounds of speech are absolutely and quickly identifiable because continuous variations in the acoustic stimulus are phonemic distinctions that are categorical in nature for most speech continua. The above researchers at the Haskin's Laboratories have arrived at their motor theory by first looking at the acoustic cues of speech themselves, i.e., at those aspects of the speech wave on which the identification of the phonemes depend—and then asking whether there is anything about them which might tend to make them inherently distinctive. They offer the opinion that the acoustic cues appear in and of themselves to be quite ordinary. For some phonemes the extremely distinctive difference one hears in the case of speech is considerably less distinctive, if indeed the difference can be heard at all, when the variable is listened to in isolation or in a non-speech pattern which is most nearly equivalent. This strongly suggests that distinctiveness is not inherent in the acoustic signal, but is added as a consequence of linguistic experience. More important, perhaps, it indicates that even with a considerable background of linguistic experience on the part of the listener, the acoustic signal is distinctive only when, being heard as a speech sound, it engages some kind of speech perception system.
Two questions thus arise: How is this distinctiveness increased when the incoming signals enter this speech perception system, and further, what are the properties of the system? One answer to both questions can be developed out of research findings by the Haskin's group which indicate that some of the consonants are perceived categorically. The impressionistic data, in a typical case, are to this effect:

When we listen to a series of synthetic speech sounds in which the second-formant transition is varied progressively in such a way as to produce in succession /b/, /d/, and /g/, we do not hear a gradual change corresponding to the gradually changing stimulus; rather, we hear the first 3 or 4 stimuli as identical /b/’s, then very abruptly with the next stimulus, the perception is of /d/, where it remains essentially unchanged until again abruptly, it shifts to /g/. Acuity for discrimination is found to be considerably greater across phoneme boundaries than within phoneme class. To this extent, a listener can discriminate sounds only as well as he can identify them absolutely as phonemes. Because of the discrimination peaks at the phoneme boundaries, the incoming sounds are heard categorically, i.e., in absolute terms rather than in relation to other stimuli . . . and they are, therefore, quickly and accurately sorted into appropriate phoneme "bins."

The average adult English speaker quickly sorts the phonemes into the appropriate bins as he has been over-conditioned to hear only gross discriminations between the phonemes he habitually uses; the trained or more phonetically experienced adult English speaker, however, may be aware of finer discriminations between phonemes and, therefore, be able to sort them into more distinctive bins than the average speaker. According to the Haskin's "motor
theory," it is to these phoneme bins that the adult refers when perceiving speech. As the present study dealt with children, one might ask if the same may be true for them. The writer hypothesizes that possibly children may indeed have their own phoneme bins to refer to when perceiving speech, although these bins may not be identical to the adults'. One purpose of this study, as discussed later, was to determine if this may be the case for children who participated in this study.

From an articulatory standpoint, vowels are different from stops and some of the other consonants in that the articulators can move continuously from one vowel phoneme to another. One might expect, then, that the perception of vowels would be quite different from stops. Liberman et al. (1962), have confirmed for vowels, by experimental studies, that there is no increase in discrimination at phoneme boundaries; moreover, the obtained discrimination functions lay considerably above those that were derived on the assumption that the listener can only hear these sounds phonemically, which is to say that the listener heard many intra-phonemic variations. This is to say that the perception of vowels is considered continuous, whereas the perception of stops is very nearly categorical. This continuous perception of vowels is an explanation for regional accents and cultural dialects, and the varied vowel systems within these two.

Much earlier than Liberman et al. (1962), phoneticians
such as Stetson (1957) emphasized the central role of articulation in speech perception. Following are accounts of several other authorities who also concur with the theory of speech perception postulated by the Haskin's group.

Stevens (1960):

Thus, in the synthesis process . . . a representation of the signal at the articulatory level will certainly occur . . . a similar representation may likewise exist at some stage during the reverse process of speech recognition [p. 53].

Luria (1966) states that he feels the closest participation of the articulatory apparatus is required for the development of the ability to perceive spoken sounds and to hear speech, and that it assumes its final character only in the process of active articulatory experience (p. 102). Further support is offered by Hockett (1955) who says:

We may suspect that [as Jack] listens to Jill, his Speech Receiver is able to decode the signal partly because the incoming signal is constantly compared with articulatory motions which Jack himself would have to make in order to produce an acoustically comparable signal. . . . In learning a foreign language, one has considerable difficulty hearing correctly until one can also produce correctly [p. 7].

Experiments by Prins (1963) have shown that children who tended to confuse place of articulation during speech sound production also had difficulty discriminating minimal word pairs in which a single phoneme was altered in terms of place of articulation. This finding, when related to previous psycholinguistic research, suggests that sound
discrimination ability is a function of articulation.

Chistovich (1965) says that

There are definite arguments in favor of the fact that prior to reaching a decision concerning the phonemes, man identifies the characteristics of articulation of the audible signal.

The "motor theory of speech perception" (Liberman, et al., 1963) also maintains that the categorical perception of some phonemes arises from a learned connection between speech sounds and the articulations that generate them and that "in time these articulatory movements . . . come to mediate between the incoming acoustic stimulus and its ultimate perception." This is to say that whether or not ongoing feedback is necessary for monitoring in the adult, there is something about the early experiences of a child or the combination of experience and "wired-in" mechanisms that has organized his auditory perception of speech in such a way that it is closely locked in with the way he produces speech. Luria (1966) would agree with this idea, and says that the first years of speech development are taken up with this acquisition of the ability to hear speech, with the participation of articulation. This process of auditory-articulatory analysis is at first manifest and overt in character. As electromyographic studies have shown (Sokolov, 1959; Novikova, 1955; and Locke, 1955), it recedes into the background only gradually, so that when or shortly before the child begins to attend school, the
hearing of speech ceases to require the actual participation of articulatory actions; they have become stored neural-motor commands.

There are fundamentally three kinds of evidence that have been viewed as favoring a motor theory of speech perception (Lane, 1965): (a) when diverse perceptual responses are evoked by a set of acoustically similar speech signals produced by diverse patterns of articulation; (b) when similar perceptual responses are evoked by a set of acoustically diverse speech signals produced by similar patterns of articulation; and (c) when both articulation and the acoustic signal change continuously over some range but changes in the perceptual response correspond more closely to changes in the articulation.

The "motor theory of speech perception" has, however, been attacked on several grounds. Fant (1963) tends to question the theory and states,

To me the reference to articulation serves primarily a function within the meta-language whereby we as outside observers may conveniently describe speech. But is it actually a part of speech perception? The alternate view I would like to propose here is that if the auditory analysis in the hearing process has proceeded so far as to allow the proposed articulatory matching, the decoding could proceed without an articulatory reference [p. 1].

It is in response to this idea that Abbs and Sussman (1971) have proposed a "feature detector" theory of speech perception. This view does not depend on a particular distinctive feature system, but rather concerns itself with the process
of auditory decoding of the acoustic speech signal which
results in phonetic identification.

Opposed also by Jacobson and Halle (1956), they say

The theoretically unlikely surmise of a closer rela-
tionship between perception and articulation than be-
tween perception and its immediate stimulus finds no
corroboración in experience: the kinesthetic feedback
of the listener plays a very subordinate and inciden-
tal role. Often we acquire the ability to discern
foreign phonemes by ear without having mastered their
production [p. 34].

Chomsky (1965) has regarded the phoneme superfluous
in any case, arguing that higher order grammatical and
semantic features override segmental cues. This is to say
that semantic and syntactical rules of the language predict
the phonetic realization of the language.

Moffitt (1969) has focused on the speech perception
capabilities of twenty- to twenty-four-week-old infants.
He attempted to determine whether or not an infant could
discriminate between the acoustic cues that are known to
be sufficient for the phonemes /b/ and /g/. Using heart
rate deceleration as an index of discrimination, Moffitt
showed that infants possessed the capacity to distinguish
between the synthetic speech stimuli. This evidence indi-
cates, according to Moffitt, that infants are able to dis-
riminate sounds long before articulation and without the
benefit of "matching neuromotor commands." MacNeilage and
Rootes (1967) studied a seventeen-year-old dysarthric girl
whose intelligence and hearing were normal. Despite severe
speech production deficits, speech perception approached normality, even in some characteristics which, according to the "motor theory of speech perception," are dependent on the listener's referring to normal speech motor control. Reference to normal motor information does not, therefore, appear necessary in all cases for speech perception.

Lenneberg (1962) presented a case, typical of a larger category of patients, where a neurological defect prevented the acquisition of the motor skills necessary for speaking a language but evidence was presented for the acquisition of grammatical skills as required for the complete understanding of the language. Lane (1965) offers both methodological criticisms and counter-evidence. Denes (1964) carried out an experiment to observe how far being able to listen to our own voice, and thereby getting a chance of associating our articulatory movements with the sounds produced by these movements, makes learning to recognize speech easier. To his disappointment, the tests produced no firm evidence to support the "motor theory of speech perception."

On reviewing the literature relating to the "motor theory," the writer noted that in very few cases was there any evidence given for a definite causal relationship between "production and perception." On the other hand, most studies and discussions leaned towards favoring a co-existing or correlating relationship between the two.
Liberman (1972, personal communication) stated that "the only empirical evidence for a motor theory of speech perception is that perception seems so clearly to follow articulatory parameters, not acoustic ones." He went on to say that "the theory does not say that in order to apply the 'model' the listener must speak, even covertly, or even be able to speak." But does the theory not imply this? If the theory is valid and one perceives speech in reference to articulation, how is it that a person who does not speak is able to perceive speech?

It is, then, the implication of the "motor theory" that one must articulate in order to perceive speech that forms the basis of research for this study.

STATEMENT OF THE PROBLEM

The hypothesis that the process of speech generation plays a role in the perception of speech has appeared periodically in the literature for many years. There is a vast number of published reports which either support or contradict the hypothesis that perception follows production. In fact, this body of research is more marked by its disagreement and inconclusiveness than by any kind of general trend. Remarkably, very few, if any, studies have been addressed to the question of why the research is in conflict.

When a child's phoneme production differs from the
norm for his age, the assumption is often made that something is wrong with his perception, apparently because it is thought that everything must go in correctly if it is to come out properly. So predominant has been the role of perceptual training in articulation therapy that it has even been considered by Van Riper (1958) to be "valuable in itself." This would imply that "production follows perception," which would be in opposition to Liberman et al.'s (1963) "motor theory of speech perception," which implies that "perception follows production."

It is of interest, then, whether this implication by the Haskin's Laboratories that "perception follows production" can be extended to include the child with a speech defect. It has been the writer's observation, and that of other clinicians, that there have been children in whom an organic defect or functional factors prevented the acquisition of some or all of the motor skills necessary for speaking a language, but that these same children presented evidence showing that the grammatical skills required for the complete understanding of the language had been acquired. Lenneberg (1962) cites such a case. Also, if the perception of spoken speech depends upon articulatory skills, then how can we explain the phenomenon whereby a deaf child (or adult) who has little or no expressive speech is still able to understand a great deal more than he can say? The same could be said of the child under one
year of age who has no meaningful expressive speech, yet understands much of what is said to him. It is known that "understanding precedes expression" (Gesell, 1963).

The purpose of this study was, then, to question and test the theory that "perception follows production" and to present evidence concerning the theory. This study attempted to show that a child who is unable to perform a specific articulatory skill necessary for producing a specific phoneme was, indeed, capable of perceiving and identifying that same phoneme. Specifically, an attempt was made to show that the children who participated in the study, although capable of perceiving and identifying a specific phoneme in another's speech, were unable to perceive and identify that same phoneme in their own speech. An initial identification task was designed which required each child to listen and respond to the experimenter's speech as stimuli. Considering that each child might be perceiving speech in reference to his own articulation or phoneme bin distinctive for him, a second identification task was designed using his own speech as stimuli. It was hypothesized that the child would have more success in perceiving a phoneme (one he is unable to produce correctly) in the experimenter's speech than in perceiving that same phoneme in his own speech. This being true, it would appear that the child is not perceiving phonemes in reference to his own articulation and some doubt may be cast on
the theory that "perception follows production."
Chapter 2

PROCEDURE

SUBJECTS AND SELECTION CRITERIA

The subjects consisted of 10 children, 9 males and 1 female, chosen on the basis of the following criteria:

1. having minimum measured receptive vocabulary score of 90 and a maximum of 120 as determined by the Peabody Picture Vocabulary Test, which would be considered within the range of normal.¹

2. hearing within normal limits.

3. being 6 to 8.6 years of age.¹

4. all subjects had General American accents.

5. all subjects presenting a diagnosed²

¹See Table 1 for Peabody Picture Vocabulary Test scores and ages of the subjects.

²Each child was diagnosed using the Templin Parley Articulation Screening Test, and the experimenter's subjective evaluation of each subject's articulation in spontaneous speech. Each child was told the illustrated Bus Story (Catherine Renfrew, F.C.S.T., 1969) and was then asked to retell the story to the experimenter using his/her own words and thus the child's spontaneous speech could be subjectively evaluated. In order to ensure that a /w/ was used for /r/ rather than a distortion of /r/, each child's tape of his/her own articulation of the stimulus words used in the identification test was listened to by another graduate student, and she was required to write down the words she heard; if she thought that all the words had
functional\(^3\) articulation defect involving a consistent substitution, rather than distortion, of \(w/r\) prevocally. Each subject was unable to produce the /\(r/\) phoneme correctly in isolation by direct imitation of the experimenter. Each subject had no more than four articulation defects including the defective /\(r/\) sound.

6. all subjects were given the Goldman-Fristoe-Woodcock Test of Auditory Discrimination in order to screen out those who may have a central auditory deficit.

Initially, all subjects were tested at 500, 1000, and 2000 Hz binaurally on the Allison 22 audiometer to determine if hearing thresholds were within normal limits for these frequencies. (Normal was considered to be anything better than 20 dB ISO.)

The Templin Barley Articulation Screening Test was then administered to each subject. From this each child's phonological system was determined; specifically, whether the child substituted \(w/r/\). Spontaneous speech was evaluated begun with the /\(w/\) sound and she heard no /\(r/\) sounds, then the child's defect was considered to be a /\(w/\) substitution rather than a distortion of /\(r/\), and the child was thus accepted as a valid subject for the study.

\(^3\)Functional here is taken to mean that there is no evidence of any organic pathology.
also. The subject was then asked to produce an /r/ by direct imitation of the experimenter. If he was unable to do this correctly and considering the other criteria, he was used in the study.

The Goldman-Fristoe-Woodcock Test of Auditory Discrimination was then given each subject so that those with a central auditory deficit could be eliminated. The criterion used was that each subject must score the norm or above for his age level.

Table 1

P.P.V.T. SCORES AND AGES OF SUBJECTS

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Age</th>
<th>P.P.V.T.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.4 yrs.</td>
<td>109</td>
</tr>
<tr>
<td>2</td>
<td>6.11 yrs.</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>7.7 yrs.</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>6.9 yrs.</td>
<td>102</td>
</tr>
<tr>
<td>5</td>
<td>6.4 yrs.</td>
<td>114</td>
</tr>
<tr>
<td>6</td>
<td>7.9 yrs.</td>
<td>95</td>
</tr>
<tr>
<td>7</td>
<td>6.6 yrs.</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>6.3 yrs.</td>
<td>111</td>
</tr>
<tr>
<td>9</td>
<td>7.2 yrs.</td>
<td>119</td>
</tr>
<tr>
<td>10</td>
<td>7.6 yrs.</td>
<td>95</td>
</tr>
</tbody>
</table>
MATERIALS USED

All hearing and discrimination testing was done in an IAC Test Suite and an Allison 22 Audiometer was used, as well as a Viking Tape Deck. These tests were administered via TDH-39 Earphones. The articulation tests and Peabody Test were given in a therapy room at a small table with the subject and experimenter facing each other.

A two-part test was designed by the experimenter in order to determine each subject's ability to identify the /r/ phoneme in both the experimenter's and the subject's own speech. The test consisted of 40 4 x 8 white cardboard cards with two black outline pictures (separated by a heavy black line) illustrating the stimulus words. There were ten pairs of rhyming words illustrated, the only difference being the initial consonant. Each card was duplicated four times. One of the word pair began with a /w/, while the other word began with an /r/, e.g., wing, ring. (See Appendix A.) Each rhyming word was placed in the carrier phrase "Point to..." This carrier phrase was used in order to serve as an alerting phrase and also as a natural context before the stimulus words. The stimulus words were first tape recorded by the experimenter and then by each subject. Both tapes were presented to each subject along with the illustrated stimulus words. The tapes were presented binaurally through TDH-39 Earphones at 70 dB hearing...
level. The tapes were made on a Rheem Califone AV Series Solid State 70-TC tape recorder at 7½ ips using RCA Magnetic Tape. They were presented using a Viking Tape Deck and Allison 22 Audiometer calibrated to ISO standards. Testing was done in an IAC Test Suite.

A standard set of instructions was given to each child for making his tape and for taking the identification test. (See Appendix A.)

TESTING PROCEDURES

Each subject participated individually in the experiment. (One subject was recorded and tested at a time.) Subjects participated either after school or on the weekend; the entire screening and testing session took between two and two and one-half hours, including several rest periods.

The experimenter had made her tape before running subjects. Twenty stimulus words were recorded twice (making a total of forty stimulus words) in the carrier phrase "Point to. . . ." with a silent interval of 8-10 seconds between stimuli. The /r/ and /w/ words were recorded in a randomized order. The tape lasted about 8 minutes. The tape was then judged by another graduate student who was asked to write down the stimulus words she heard in order to ensure that the tape was of adequate quality for the subjects to be able to identify specific pictures representing the stimulus words.
After screening each subject for hearing, vocabulary, and discrimination ability, a 15-minute rest period was taken before making the child's tape.

The experimenter and child entered an IAC Test Suite (subject's side) and the subject was seated on a chair facing the microphone, which was placed on a table at the child's mouth level. The experimenter sat beside the subject with the tape recorder placed on the floor next to her so that she could easily manipulate the pause control. In making the child's tape he was instructed to mimic the experimenter saying the phrase "Point to..." The pause control on the tape deck was used so that only the child's voice was recorded. Again, an 8-10 second silent interval was used between stimuli. The words were recorded in a different order than the experimenter's. Each tape was about 10-12 minutes long. After making the child's tape, another short rest period was taken before giving the identification tasks. The experimenter and another graduate student (the one judging the child's tape to ensure that /w/ was an /r/ substitution rather than a distortion of /r/) listened to the tape during this time to see that it was of a quality good enough for the test, i.e., they were able to understand what was said.

The experimenter and subject again entered the IAC Test Suite and the child was seated opposite and facing the experimenter. The child was shown the pictures
representing the stimulus words. A training session was carried out so that the child would know what the pictures were. The first set of ten cards, i.e., all twenty words, was used. The child was told what each picture was and was then asked to point to each one separately. If the experimenter felt that the child was unsure of the pictures after going through the first set, she went through the second set (the same twenty words) in the same manner. All subjects were easily able to recognize the pictures after the training session. Instructions were then given for the actual test. (See Appendix A.)

Headphones were then placed on the child and the experimenter's tape was turned on. The experimenter turned the cards as the child responded by pointing to one of the two pictures. His responses were recorded on a score sheet. (See Appendix B.) After listening to the experimenter's tape, the headphones were removed and the child was given a short rest while the child's own tape was put on the tape deck. The instructions for the test were repeated (see Appendix A), and the headphones were then put back on and the child's tape was turned on. Again, the experimenter held and turned the cards and recorded the child's responses (see score sheet, Appendix B). As the experimenter's and the child's stimulus words were recorded in different orders, the experimenter had to turn the cards in a different order for each tape.
ANALYSIS OF DATA

It was hypothesized that the results of the first part of the identification test, using the experimenter's speech, are going to show a positive outcome with not less than 90% correct responses for each subject in identifying a phoneme which he is unable to produce. This percentage score is far greater than the "motor theory of speech perception" would allow for. Each subject with a score of 90% or greater would tend to support the writer's hypothesis that a child does not have to be able to produce a phoneme correctly in order to identify that same sound correctly in another's speech. The writer also hypothesized that the results of the second part of the test, using the child's own articulation, would show an outcome of not more than 50% correct responses for each subject in identifying a phoneme which he is unable to produce correctly. This finding would also tend to oppose the "motor theory" as it would seem that the child was not perceiving sounds in reference to his own articulation system either.
Chapter 3

RESULTS

The data obtained in this study consisted of four separate scores for each subject; the first being the score made by each subject after listening to the experimenter's tape, and the second being the score obtained after listening to the subject's own tape. This score can then be further evaluated by looking at both the score obtained for the /r/ words and the score obtained for the /w/ words.

As can be seen from Figure 1, all results for each subject are remarkably consistent. All subjects scored 100% correct responses after listening to the experimenter's tape. This confirms the writer's hypothesis that each subject would score 90% or better on listening to the experimenter's tape. Each subject scored only 50% correct responses on listening to his own tape. This again supports the writer's hypothesis that none of the subjects would score over 50% correct responses (for the sound he was unable to produce) on listening to his own tape. Looking more closely, the 50% correct responses were all the /w/ words; no subject got any of the /r/ words correct on listening to his own tape. When hearing his own
Figure 1. Scores for all 10 subjects for both tapes and number correct for /r/ and /w/ words for both tapes.
articulation of an /r/ word, each subject responded by pointing to its rhyming /w/ word cognate.

Although the subjects who participated in this study were only a sample of the population, the results were deemed significant in that all subjects performed in exactly the same manner. Evidence against the validity of the "motor theory" would have been seen if only one subject had performed as hypothesized, but the fact that ten subjects performed as hypothesized provided very strong and significant evidence in disfavor of the "motor theory" for this sample. The writer feels that in light of these results the probability is great that other children in this "population" will perform in the same manner.

---

Population here was taken to mean children between the ages of 6 to 8 years with normal hearing and I.Q. and who exhibited functional articulation defects.
The purpose of this study was an attempt to provide some evidence either leaning for or against the motor theory of speech perception which states that "perception follows production." In particular, 10 children who were unable to correctly articulate /r/ and substituted /w/ for it were presented with tapes made by both the experimenter and the subjects themselves which consisted of paired rhyming words, one beginning with /w/ and the other beginning with /r/. Each child was required, while listening to the tapes, to respond by pointing to which illustrated stimulus word they heard. The subject's responses were scored.

The results showed that each child scored 100% correct responses while listening to the experimenter's tape and only 50% correct responses while listening to his own tape (all correct responses being /w/ words). These results would tend to cast some doubt on the validity of the motor theory of speech perception: "perception follows production" for these cases. The doubt is found in the observation that each child was perceiving, in the experimenter's tape, the phoneme /r/ which he was unable to correctly produce by direct imitation, and it was felt therefore that
he could not be referring to his own articulation of that sound. It would be a mistake to consider the aforementioned finding as conclusive evidence, and it was for this reason that the child was presented with his own articulation of the words used for the experimenter's tape and his responses scored. If it were the case that each child had his own "phoneme bin" that he referred to for the identification of the /r/ sound, then it would be reasonable to assume that each child would be able to perceive correctly the /r/ sound when he listened to his own speech. This was not the case, and no child correctly responded to the /r/ words, implying that they did not have a specific phoneme bin for /r/ to employ for articulatory reference. If the child is not referring to his own articulation to identify the /r/ sound, then additional doubt can be cast on the validity of the motor theory in considering children with a functional articulation defect involving a consistent substitution of one sound for another, in this case w/r. In arriving at their motor theory, the Haskin's group experimented only with adult college subjects; no known Haskin's studies researching speech identification and discrimination have been carried out using normal or speech-defective children as subjects. The writer feels that this indeed would be an area of study where further research would be interesting and beneficial in learning more about the role of perception, identification, and discrimination.
abilities and their relationship to each other in children's acquisition of speech and language skills.

The relationship of the ability to distinguish sounds uttered by someone else (external model discrimination) to the ability to monitor one's own sound production errors and the relation of both to articulatory proficiency is a topic which has been mentioned in discussions for many years, but has not been studied in any real systematic sense. The writer did find, however, a study by L. F. Aungst and J. V. Frick (1964) designed to investigate the hypothesis that, for children 8 years of age or older and who misarticulate the /r/ sound only, consistency of articulation is more directly related to the ability to judge one's own speech productions as correct or incorrect, than to the ability to discriminate between paired auditory stimuli presented by another speaker. Their findings concur with the present writer's in that it is felt that the ability to articulate /r/ may not be related to the ability to discriminate between paired "external" auditory stimuli presented by another speaker. However, the ability to articulate /r/ seems to be related to the ability to judge the "correctness" of one's own speech productions under conditions of instantaneous judgment. These findings lend support to the importance of self-monitoring ability in adequate speech production.

The fact that this study was concerned with
identification, discrimination, and articulation relationships creates the question of therapeutic implications; specifically, the role of ear training in functional articulation cases. One of the first skills many speech clinicians teach children with misarticulations is to discriminate between correct and incorrect sounds. According to a national survey, clinicians devote a considerable amount of time to this task (Chapman et al., 1961). Training in auditory discrimination has long been deemed a necessary prerequisite to learning articulation (Powers, 1957). Van Riper (1947, p. 173) stated, "It may be said with the utmost emphasis that no teacher should attempt to get a child to try and make a new sound without first giving him systematic ear training." He went on to say that after intensive ear training the child often may produce the correct sound on his first attempt. He does not mention, however, that the child may be capable of producing the sound correctly even before systematic ear training. Still today Van Riper (1972) states that "in the first phase of therapy for articulation cases, the emphasis is all on listening. It is ear training."

If discrimination training is viewed as an example of response generalization within the positive transfer model, it would seem entirely possible that learning Task A (discrimination) might facilitate learning Task B (producing the sound). Whether or not this is the actual
case seems to warrant further research in this area.

Holland and Matthews (1963) provide some helpful insight into this problem. Three different sound discrimination programs were administered to three groups of children. Each group consisted of nine children judged to have defective /s/ articulation. A battery of tests designed to evaluate sound discrimination skills and ability to produce /s/ was given before and after discrimination training. After training, children in all groups improved their /s/ articulation scores. This would indicate at first glance that response generalization occurred, that is, learning to discriminate (response$_1$) facilitated learning to produce new sounds (response$_2$), implying that "production follows perception" in these cases. However, the authors cautioned that improved articulation ability did not indicate that children who could not produce /s/ before the discrimination training could produce /s/ after. It simply meant that some children were able to articulate /s/ correctly more often.

Of particular interest in the present study and in Holland and Matthews' (1963) was the observation that the children frequently vocalized aloud or whispered many of the /r/ (/s/) words during the training period and again during the actual test. In the present study the children articulated the /w/ sound only for the /r/ words (the experimenter paid very close attention to this phenomenon
as she felt it relevant to the study). In the Holland and Matthews' study the children were noted to articulate both correct and incorrect /s/ sounds. Holland and Matthews suggested that these unsolicited vocalizations may have accounted for the improvement in /s/ articulation scores by forcing the children to carefully observe and rehearse certain auditory cues. They feel that in this case correct or incorrect discrimination may be an irrelevant task, whereas saying /s/ may be the important feature of discrimination training, i.e., "perception follows production."

The present writer feels that she could not conclusively say that the child's unsolicited vocalizations may be responsible for helping the child decide which stimulus word was said. This is felt to be so due to the fact that the child, when listening to the experimenter's tape, would hear an /r/ word, e.g., "ring," and would vocalize "wing" but still correctly point to "ring." However, when listening to his own tape he would hear his own production of "ring," i.e., "wing," and vocalize "wing" and then point to "wing." This inconsistency again raises the question: is defective articulation the cause of auditory discrimination defects or vice versa?

Stitt and Huntington (1969) have presented an extensive review of the literature concerning the relationship between auditory discrimination and incidence of misarticulation. They justify continued emphasis upon auditory
distraining in speech therapy solely on the basis of correlation between poor discrimination ability and poor articulation. However, a correlation does not guarantee a causal relationship.

Locke (1968) suggested that it may be unwise to spend time teaching interdiscrimination skills, that is, the ability to make discriminations of someone else's sound productions. He pointed out that it might be more helpful if the clinicians make cues distinctive for the children by pairing the correct sound with an already distinctive visual stimulus such as a picture of an object or animal. The problem is to determine which activities provide positive transfer and which do not.

After studying the results of the present study, the writer is of the opinion that one cannot emphatically say that "perception follows production," nor that "production follows perception" is the only "model" operating for a child's ability to correctly perceive or produce speech. It would appear to the writer that the two "models" may not be considered as operating separately, but that they are integrated and dependent upon one another. However, it may occur that one "model" may emerge in a child before the other or that both "models" emerge at the same time. This is thought by the present writer to be a possible explanation for discrepancies between children's articulation and discrimination abilities, and this will be
discussed in relation to the present study. All the children were able to perceive correctly all the /r/ words in the experimenter's speech, and in this instance it would appear that perception had preceded production. One must not stop here but consider that the children could not perceive the /r/ sound in their own speech. The writer is of the opinion that the children will not be able to accomplish this successfully until after they are able to make an /r/ sound or some perceivable distortion of it so that they have a specific phoneme bin for the /r/ sound and thus something to refer to when attempting to perceive the /r/ in their own speech. Therefore, in this instance it could be said that "perception follows production." Each child is of course an individual, and should be considered as such when looking at speech development. The above hypothesis may explain why it is that some children have little or no trouble discriminating a sound in others' speech while they cannot produce it, while another child may be well able to produce the sound while being unable to discriminate it. In each case only one "model," i.e., "perception follows production" or "production follows perception," may be working. For the normal child, that is, one with no articulation or discrimination problems, it may be that both "models" are working effectively. If the hypothesis were true, then further research concerned with children's misarticulations and auditory
discrimination abilities might arrive at some pattern where it can be seen that articulation problems are either the result or cause of discrimination problems. This knowledge could then be used for planning more effective therapy programs for each child. The writer feels that in view of the preceding hypothesis both "models" must be working properly in order to achieve correct articulation and discrimination skills.

It is felt that the speech clinician cannot truly separate ear training and articulation training (as has been the practice of several authorities) when planning a therapy program for these children. It would seem that initially one type of training should not take precedent over, to the exclusion of, the other, but both should be combined in therapy. For example, when looking at the children in the present study, it would seem reasonable to use ear training in teaching them to perceive and discriminate between the experimenter's production and their own production of the /r/ sound while concurrently using articulation training in an attempt to teach them the correct articulation of the /r/ sound after which further ear training will have to be incorporated into the therapy program in order to teach the child to perceive and discriminate between his own correct and faulty articulations of the /r/ sound. It is felt by the writer, then, that it is of utmost importance not to concentrate strictly on one
aspect of therapy, for most children, but to combine and integrate several aspects, i.e., ear training, articulation, and discrimination, among others.

Although only one female was used in the study, and her results were consistent with those of the males, further research using more females might show some discrepancies in results.

RECOMMENDATIONS

It is evident from this study and from the literature that there is a good deal of discrepancy of thought in the areas of speech perception, production, and discrimination and their relationships to each other. For this reason it is obvious that more research is needed in this field. It is the writer's opinion that there may possibly be studies which could be carried out in relation to the present one which might provide further knowledge to this area. Following are some of the hypothetical possibilities:

The writer feels that it might be of interest to take preschool children at different age levels who have a w/r substitution and devise an identification task somewhat like the one presented in this study, but possibly more suited to younger children, in an attempt to determine at what age, if any, the child is unable not only to perceive the /r/ in his own speech but also in the experimenter's speech. Results from studies such as this might
provide some normative data which could be used in the assessment of speech and discrimination development in children with speech and discrimination problems.

Not only might the above line of investigation be interesting, but it is felt that it would also be interesting to do follow-up studies of the children used in the present study at different intervals to see if there is a mean age or point along the /w/ to /r/ continuum where the child would be able to perceive the /r/ phoneme in his own speech, thus implying that an /r/ phoneme bin has been established for the child to refer to, if such is the case. This phoneme bin may not be exactly like that of the adult's but would be distinctive for each child. This type of study would be almost impossible for the average clinician researcher to carry out unless he had access to a spectrogram or some other mechanical method for judging the acoustical properties of the sound produced in order to determine where along the /w/ to /r/ continuum the child's /r/ sound lay. This might then be an area of investigation for a group such as Haskin's. They would be able to synthetically produce phonemes along the /w/ to /r/ continuum and design tasks for both normal and speech-defective children (problem with /r/ in this case) to determine at what point along the continuum they perceive /r/ as distinct from /w/, if at all. The point where the child perceived /r/ could be compared with his own production of /r/, and
evidence either for or against the "motor theory" could be gained for the case of the speech-defective as well as normal child.

One study similar to the one hypothesized above was carried out by P. Menyuk and S. Anderson (1969). The purpose of their study was to examine preschool children's identification and reproduction of the speech sounds /w/, /r/, and /l/ and to compare the performance of children and adults in these tasks. The stimuli consisted of three sets of synthetically produced CVC syllables that ranged in equally spaced formant contour changes from "light" to "white," "light" to "write," and "white" to "write." Subjects were asked to reproduce the word they heard, and to identify it by pressing a button under a picture of the word. Neither children nor adults observed sharp boundaries between the speech sounds in this set. The responses of children were different in the reproduction and identification tasks. More children observed speech sound boundaries in the identification than in the repetition task, and significantly more frequently produced /w/ in response to the stimuli than the other two sounds, but they did not identify /w/ significantly more often. These results were not found with the adult population. The authors of this study hypothesized that the developmental sequence in the acquisition of the members of this set is, first the ability to identify differences between members of the set,
and second to reproduce the differences. They therefore found that "production follows perception" for these cases.

It does appear that further research in the areas of speech perception, production, and discrimination is needed so that more insight can be gained concerning these topics and thus more effective methods of diagnosis, assessment, and treatment for children with speech problems can be determined.
Chapter 5

SUMMARY AND CONCLUSIONS

The purpose of this study was to provide evidence concerning the motor theory of speech perception which states that "perception follows production." In particular, 10 children, 9 males and 1 female, between the ages of 6.3 and 7.9 years of age participated in the study. All subjects were unable to correctly produce the /r/ sound and substituted the /w/ sound for it. They were presented, individually, with tapes made by both the experimenter and the subjects themselves. Each tape consisted of 40 paired rhyming words (which were illustrated on cardboard cards), one word beginning with /r/ and the other beginning with /w/. Each child was required, while listening to the tapes, to respond by pointing to the illustrated stimulus word they heard. The subjects' responses were scored.

The results of the experiment did not support the theory that "perception follows production" for the children in this study, but the writer felt that the theory was not irrelevant in a theory of speech perception, and reasons for this were discussed. It was felt that possibly the motor theory of speech perception may be one "model"
working in a broader theory of speech perception for children with functional speech and discrimination problems.

Other studies related to the present study were mentioned, and consistencies and discrepancies between these and the present study were discussed. Implications for further research were presented, along with some studies already carried out in relation to the implications.
LIST OF REFERENCES


40

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
APPENDIX A

Instructions given by experimenter to each child for:

A. Making child's tape.
B. Training session for cards.
C. Taking first part of identification test (experimenter's tape).
D. Taking second part of identification test (subject's own tape).

A. (child's name) ______, we're going to make a tape of you saying some sentences. Turn around so that you are speaking into the microphone. After you make the tape, then I'll let you listen to it through these headphones in a little while, OK? I want you to say this after me: "Point to ring," etc. Good, now let's turn on the tape and when I point to you, you say the sentence I just told you, then be very quiet and don't talk or make any noises until I tell you to say the next sentence. Any questions? Good. (Turn on tape and point to child; child says phrase "Point to . . . ." and is quiet for 10 seconds after which the pause control is pushed and the tape is stopped.) Now say this when I point to you: "Point to . . . ." (Continue in
this manner until the entire tape is made.) Good work; now let's take a rest before we listen to and look at some pictures.

B. (Experimenter shows child cards with stimulus pictures on them.) See these pictures? This is a dog wagging his tail, and this is an old piece of cloth called a rag, etc. Now listen carefully because I'm going to say one, and I want you to point to the one I say. Point to rag, wag, etc. (All the cards in the first set were gone through in this manner.)

C. Good, now that you know what all the pictures are, I'm going to put these headphones on you and you'll hear me on a tape telling you which picture to point to. You have to listen very carefully. If you're not sure what word I say, then guess. It's better to guess than to not point at all. Any questions? Good; I'll put the headphones on now and you listen and point carefully. (Turn tape on and record child's responses.) Well done; we'll take a short test now before we put the headphones back on and you listen to the tape you made.
Test 1

(Time interval 8-10 seconds between stimuli)

1. Point to wag
2. Point to rip
3. Point to rake
4. Point to one
5. Point to rich
6. Point to rock
7. Point to whale
8. Point to wheel
9. Point to wing
10. Point to rain
11. Point to whip
12. Point to wake
13. Point to run
14. Point to witch
15. Point to walk
16. Point to rag
17. Point to reel
18. Point to ring
19. Point to Wayne
20. Point to rail

21. Point to rake
22. Point to one
23. Point to whale
24. Point to rain
25. Point to wag
26. Point to wheel
27. Point to rip
28. Point to rock
29. Point to wing
30. Point to rich
31. Point to ring
32. Point to walk
33. Point to witch
34. Point to rail
35. Point to run
36. Point to Wayne
37. Point to whip
38. Point to reel
39. Point to wake
40. Point to rag

D. Now you're going to listen to your own tape telling you what pictures to point to. Listen very carefully and guess if you're not sure. Any questions? OK, here
we go again. (Put headphones on, turn on tape, and record child's responses.)

**Test 2**
(Time interval 8-10 seconds between stimuli)

1. Point to ring
2. Point to walk
3. Point to witch
4. Point to rail
5. Point to run
6. Point to Wayne
7. Point to whip
8. Point to reel
9. Point to wake
10. Point to rag
11. Point to rake
12. Point to one
13. Point to whale
14. Point to rain
15. Point to wag
16. Point to wheel
17. Point to rip
18. Point to rock
19. Point to wing
20. Point to rich
21. Point to whip
22. Point to wake
23. Point to run
24. Point to witch
25. Point to walk
26. Point to rag
27. Point to reel
28. Point to ring
29. Point to Wayne
30. Point to rail
31. Point to wag
32. Point to rip
33. Point to rake
34. Point to one
35. Point to rich
36. Point to rock
37. Point to whale
38. Point to wheel
39. Point to wing
40. Point to rain
## APPENDIX B

### Score Sheet

<table>
<thead>
<tr>
<th>Stimulus Words</th>
<th>E*</th>
<th>S*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set 1 E.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. wag 31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. rip 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. rake 33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. one 34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. rich 35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. rock 36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. whale 37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. wheel 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. wing 39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. rain 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Set 4 S.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Set 2 E.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Set 3 E.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Set 3 S.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Set 4 E.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Set 1 S.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Experimenter = E (saying stimulus words)*

*Subject = S (saying stimulus words)*

+ = Correct response   
- = Incorrect response