SPEECH HANDBOOK

Precision Programs Especially for the Hearing-Impaired

PART 2 of the Listening and Speech Package (LAS-PAC)

FREDERICK S. BERG

Part 2 of the Listening and Speech Training Package (LAS-PAC)

Frederick S. Berg, Ph.D.

Professor
Department of Communicative Disorders
Utah State University
Logan, Utah

Grune & Stratton
A Subsidiary of Harcourt Brace Jovanovich, Publishers
New York San Francisco London

Library of Congress Cataloging in Publication Data

Berg, Frederick S.
Speech handbook.

(*His* Listening and speech package; v.2)
Bibliography
Includes index.
1. Deaf—Means of communication. I. Title.
HV2471.B47 vol. 2 362.4'2s [362.4'2] 78-16395
ISBN 0-8089-1129-5

Grune & Stratton, Inc.
111 Fifth Avenue
New York, New York 10003

Distributed in the United Kingdom by
Academic Press, Inc. (London) Ltd.
24/28 Oval Road, London NW 1

Library of Congress Catalog Number 78-16395
International Standard Book Number 0-8089-1129-5

Printed in the United States of America

ACKNOWLEDGMENTS

Credit is due many individuals and agencies who aided in the development of the *Speech Handbook*. My talented wife, Edna Berg, prepared all of the illustrations. Target formats of the recording forms were drawn by Ronald Hammond, a gifted student. Lanus Thompson did an excellent job of typing the manuscript as well as many of the figures.

Needed consultation and editorial guidance were provided by Larry Bradford and by my publisher, Grune & Stratton.

The Kellogg Foundation and the Social and Rehabilitation Services supported the Listening and Speech (LAS) Project, within which the handbooks were developed. Administrative, technical, and consultative assistance was provided by personnel of the Quality of Rural Life Program, the College of Education, the Department of Communicative Disorders, and the photographic and printing services of Utah State University. Many professionals throughout the country, particularly within an eight-state region including Utah, have field-tested the programs in the *Speech Handbook*.

My last debt is owed to the hearing-impaired clients whose desire to improve their speech skills made this formidable and exhaustive effort worthwhile.

Author's note: *Home-study course work in LAS-PAC, phonetics, and educational audiology is available through the Independent Study Division of Utah State University in Logan. Workshops in these professional areas may also be arranged by contacting the U.S.U. Department of Communicative Disorders.* Educational Audiology: Hearing and Speech Management, *published in 1976 by Grune & Stratton, provides additional background information for use with the* Speech Handbook *and* Listening Handbook.

PREFACE

Speech may be defined as vocal, prosodic, and articulatory behavior. Listening is perception of speech by use of audition, vision, and a combination of auditory and visual speech clues. The *Speech Handbook* and companion *Listening Handbook* comprise the basic parts of the newly developed Listening and Speech Package (LAS-PAC). The LAS-PAC, which has been designed to extend and expand the spoken communication skills of hearing-impaired children and adults, is a powerful clinical tool in the hands of the knowledgeable and skilled clinician. It also provides practical and technically advanced text material for professional preparation.

The *Speech Handbook* includes background material, training guidelines, and stimuli and recording forms for a comprehensive core of data-based continuous-evaluation vocal, prosodic, and articulatory pgrams. Through an agreement with the author, the publisher grants clinicians permission to duplicate copies of the stimuli and recording forms for use with any number of clients. Copies for each client may be hole-punched and inserted in a looseleaf notebook and thereby utilized for individualized speech management. If a client has a minimal number of speech errors, duplicated copies of needed stimuli and recording forms from the *Speech Handbook* may also be kept in the listening workbook of that client. For a client who has many speech errors, it might be useful to keep both a speech notebook and a listening notebook. Supplementary speech management material, clinical plans, and progress reports may be kept in the speech notebook also.

Copies of forms or reports can be duplicated for clinical filing, professional referrals, and home-training use. The notebook could then become the property of the client when he or she moves to another location where another clinician would be able to utilize the information it contains for a continuation of LAS-PAC management and purchase copies of the *Speech Handbook* and the *Listening Handbook* for procedural assistance and additional stimuli and recording forms.

The critical need to upgrade the speech skills of many hearing-impaired children and adults provided the basis for the development of the *Speech Handbook*. These clients exhibit more stress, voice-quality, pitch, intonation, oral–nasal, and articulatory errors than the vocal tract and articulatory deviations often manifested by clients who have normal hearing. The *Speech Handbook* includes testing and training materials designed especially for hearing-impaired clients; each handbook program includes at least three stages, and many have five stages. The basic programs provide preliminary evaluation, sensory testing and training, and shaping and refinement of target behaviors. Complete programs include these stages, as well as transfer training and generalization testing.

Preliminary evaluation enables the clinician to identify and describe the vocal, prosodic, and articulatory skills and deficiencies that may be present in the speech of a hearing-impaired client. The *Speech Handbook* includes a brief vocal and prosodic test and an extensive articulation test. Each test item is incorporated in a basic program target with space so that the clinician can mark accuracy of response on a continuum. Initially the entire tests are administered to a client, and comprehensive data are recorded. These

CONTENTS

Speech is highly desirable behavior in our culture. It is our main vehicle for expressing our thoughts and feelings to other people. Ordinarily, people learn to speak early in life and articulate completely by 8 years of age. When individuals hear imperfectly, they characteristically speak defectively and inarticulately. Many persons without hearing loss also exhibit speech problems into adulthood. A speech problem may interfere with communication, draw attention to itself, or cause its possessor to be maladjusted (Van Riper, 1954).

Speech Parameters

Speech behavior occurs within the vocal tract, originating in the thoracic cavity and terminating at the front of the face (Berg, 1976). Classically it has been described as requiring respiration, phonation, resonance, and articulation processes. Speech is particularly dependent upon the phenomena of stress, vocal fold vibration, pitch, intonation, oral–nasal ratio, and articulation.

Stress is a prosodic phenomenon of relative durations and loudnesses of successive syllables of speech. The accents of the syllables of speech are notated by oblong marks. The marks are drawn below syllables to convey four stress variations. The length and height of each mark reveal duration and force (perceived as loudness), respectively. The relative sizes suggest primary, secondary, tertiary, and quaternary levels of stress. The first two sizes are stressed syllables, and the last two are unstressed syllables.

It is easy to distinguish between stressed (first two sizes) and unstressed (last two sizes) syllables, relatively easy to distinguish between primary and secondary stress, but difficult to distinguish between tertiary and quaternary stress. The varying patterns of speech we have developed are evidence that we perceived all four distinctions early in life while learning our spoken language. The intelligibility or meaningfulness of speech is based, in part, upon perception and production of stress patterns. Incorporated into language, these prosodic features are called suprasegmental stress phonemes.

Vocal fold vibration is a vocal phenomenon of the larynx that occurs during the production of the voiced sounds of speech. The voice is pleasant when the vocal folds vibrate precisely. Vocal disorders occur when the folds are not coming together or separating on target. The quality or timbre of voice is largely determined by the "tuning" of the vocal folds. When we call a voice *breathy* or *harsh,* we are perceiving distortion in this vibrational system.

Pitch is also a vocal phenomenon. It is our perception of the fundamental rate or frequency of vocal fold vibration. When a speech sound is vocalized, it has a pitch. During connected speech, a range of pitches occur. People have different pitches because their vocal folds vary in length. Pitch or vibrational rate also depends upon vocal fold tension. Tension must be precisely adjusted or pitch will be inappropriate. When a pitch does not match anatomical and physiological capabilities, the vocal folds fatigue and pitch is distorted. Falsetto is an example of profound pitch distortion. The pitches that people habitually use should be those they naturally use. Sophisticated instruments enable the laryngologist and speech pathologist to predict a client's natural pitch range.

Intonation is a prosodic phenomenon within the vocal phenomenon of *pitch*. Pike (1945) observed that each phrase includes an intonation contour; that is, a sequence of contrastive or relative pitch changes associated with each fully stressed syllable. Some intonation contours carry meaning distinct from those communicated by the sequence of words included. Other contours do not. Linguists have identified four distinctive pitch variations within phrases and three between them. These regular and terminal pitch phonemes combine with four stress phonemes to constitute the suprasegments of phonology.

Illustrations of various intonation and stress marking systems are included in the book entitled *Educational Audiology* (Berg, 1976). The author favors using oblong patterns to mark stress and a continuous line to mark intonation. The line varies in elevation to depict pitch changes and extends from the beginning to the end of a phrase or sentence. This prosodic marking system is also used in the LAS-PAC.

Woodward's (1967) experience at the Central Institute for the Deaf led her to believe that even profoundly hearing-impaired children learn to produce four pitch levels and associate them with intonation markings. Woodward also suggested that once such children develop flexible control of pitch, they seem to be able to produce nuances such as excitement and disgust.

Precise perception and production of intonation ordinarily depend upon refined pitch discrimination. Only an intact cochlea or perception of electrovisual clues of vocal fold vibration provides sufficient sensory input for shaping and refining intonation. Kinesthesia is grossly helpful, but speechreading and taction are noncontributive.

Oral–nasal ratio is a comparative measure of sound energy exiting the mouth versus that exiting the nose during speech. TONAR II, developed by Fletcher (1974), enables the clinician to obtain acoustic correlates of this comparative measure. Normal nasal energy during speech varies from 5 to 15 percent. The physical measure of nasal energy is called nasalance (Fletcher, 1976), whereas a person's perception of this same energy is called nasality. Fletcher concluded from a recent study that listeners are capable of reliably judging nasality. He also found that nasalance and nasality can be highly correlated. Training in rating speech samples with varying amounts of nasalance assists listeners in making valid and reliable judgments of nasality.

Fletcher (1974) notes that when velopharyngeal valving is defective, sound spills into the nose and severely disturbs speech output. The vowels tend to be blurred, and the consonants lose their precision and crispness. In addition, certain consonants are frequently omitted or replaced by others that do not require an oral pressure buildup in the mouth. With competing and distracting nasal resonance, verbal communication can be disrupted and personality and behavior deleteriously affected. Consequently, it is important to train the client to control the velopharyngeal musculature. In the sense of affecting vocal resonance, the oral–nasal ratio is a vocal phenomenon.

Articulation is the phonetic phenomenon of speech that occurs in the larynx, in the velopharyngeal area, and in and around the mouth. It may be subdivided into voicing, oral–nasal, and manner and place of articulation distinctions. Each consonant, vowel, and diphthong is a combination of (1) whether the vocal folds are vibrating, separated, or approximated, (2) whether the velopharyngeal port is open or closed, (3) degree of restriction of the exhaled airstream through the oral cavity, and (4) locality of articulation within the region immediately surrounding and including the mouth. Place of articulation depends upon the positions taken by the mandible, lips, and particularly parts of the tongue in relationship to the upper teeth, alveolar ridge, hard palate, and velum to complete the production of all phonemes. During speech the articulations of consonants, vowels, and diphthongs frequently overlap. An articulatory position of a sound is physiologically anticipated before it is produced.

Because of its complexity, articulation is the last of the speech phenomena to normally mature. Many subskills have to be developed before this achievement is reached (Ling, 1976). The focus of the listening and speech programs of the LAS-PAC is perception and production of consonants, vowels, and diphthongs in their many stimulus contexts. Prosodic and vocal phenomena are managed also.

Speech Training Developments

Instruction in speech may be traced to the beginning of the eighth century. In about 700 A.D., St. John of Beverly taught a deaf mute to speak (Streng et al., 1955). The precision with which this feat was accomplished is unknown. Even today, most deaf children in the United States have largely unintelligible speech. The task of teaching speech to a deaf client is formidable because each contributive vocal, prosodic, and articulatory skill has to be taught from a zero or near zero baseline. Very little of the speech of a deaf child develops naturally.

Clients with defective but intelligible speech typically are not deaf. They may be hard of hearing or have speech problems associated with cleft palate, cerebral palsy, mental retardation, or other organic and functional involvements. All the clinical speech subpopulations include cases with varying extents of vocal tract problems that require specialized assistance.

The technology of speech instruction has developed largely within the communicative disorders professions of education of the deaf, speech pathology, and educational audiology. During the past decade, many professionals have made advancing contributions in the design and implementation of speech remediation. With the newer technologies and with sufficient programming, specialists can assist clients with vocal, prosodic, and articulatory problems to make dramatic improvements.

Methodology in Deaf Education

Calvert and Silverman (1975) describe three major methodologies of speech instruction in the current education of hearing-impaired clients. These are the Auditory Global, the Multisensory Syllable Unit, and the Association Phoneme Unit methods. In the Auditory Global method, the teacher capitalizes on the speech skills that have developed from auditory stimulation in natural situations and in listening training. Basic vocal and prosodic competencies and considerable articulatory skill have been acquired prior to the initiation of speech training. Such clients demonstrate that they are hard-of-hearing children rather than deaf youngsters.

The Multisensory Syllable Unit method is the typical approach utilized in the speech management of deaf children. Emphasis is placed on utilizing speechreading, tactile, kinesthetic, and auditory clues and visual speech aids to stimulate the client. Programming focuses on the production of the various syllables, particularly in contextual and prosodic speech. Within this program, some specialists have focused on precision of speech and others on naturalness of communication.

The third methodology of speech management for hearing-impaired children has been the Association Phoneme Unit method developed by McGinnis (1963). Phonemes are shaped in isolation and then included in syllables, words, phrases, and sentences. In addition, children may be prompted to position their articulators before producing a given phoneme. The unisensory stimuli of speechreading clues or of auditory clues are also associated with the articulation and meaning of a given word. A core of 50 to 100 words, a series of simple sentence forms, and a sequence of more complex language forms are included. Such a structure may be the core of the academic curriculum during up to 8 years of schooling.

One of the notable speech programs in the country has evolved at the Clarke School for the Deaf since its founding in 1867. Five levels of speech instruction are utilized. During preschool years and up to age 7, emphasis is placed on encouraging children to use spontaneous speech and to develop a well-modulated voice. Thereafter, children receive intensive training to shape the consonants, vowels, and diphthongs, to associate them with phonetic symbols, and to incorporate them into connected speech. The last three levels of instruction begin when children are approximately 9 years old and end about 6 years later when they graduate from the eighth grade. During the final 6 years, the program is designed to refine vocal, prosodic, and articulatory skills and to teach the pronunciations and meanings of the children's expanding vocabularies (Magner, 1971).

Recently Ling (1976) has analyzed the skills of the formidable task of teaching speech to deaf children. He specifies subskills for many vocal, prosodic, and articulatory skills that need to be taught. He indicates that target skills must become reflexive so that needed feed-forward processes can enable these children to speak. The skills are sequenced within a seven-stage model of speech acquisition. Vocal skills are initiated in the first stage, prosodic in the second, and the many articulatory skills in the third through seventh stages. Ling also details spectrographic data for each of the consonants, vowels, and diphthongs as the basis for careful exploitation of residual hearing. In addition, he correlates phonetic with phonemic programming in his speech acquisition model. He states that the Auditory Globbal, the Multisensory Syllable, or the Association Phoneme method can be implemented within this model.

Methodology in Speech Pathology

Modern methodology within speech pathology is exemplified by the work of Mowrer (1969), McLean (1967), Bokelmann (1973), Holbrook (1972), Fletcher (1972), and Johnson (1974). The technologies of each of these innovators have been based in part upon fundamental principles of a branch of psychology called operant conditioning. The originator, Skinner (1953), viewed operant conditioning as the relationship between stimulus, response, and consequence.

Skinner (1953) discovered a three-term contingency symbolized as S^D--R--S^r. S^D is the discriminative stimulus that occasions or prompts the response, R is the response, and S^r is the reinforcing stimulus that immediately follows the response. When positive reinforcement is given, the frequency of the response increases. If negative reinforcement is provided instead, the frequency of response decreases. Neutral reinforcement results in no change of response rate or probability. Thus the consequent event contributes to acceleration or deceleration of response. The antecedent event also determines the response rate. The manipulation of the antecedent event is called stimulus control.

Mowrer and associates (1970) have developed a program format for shaping a target phoneme and transferring this correct articulation into word, sentence, and story contexts. The prototype program, called the S-Pack or S Programmed Articulation Control Kit, includes 269 verbal and picture stimuli for the /s/ phoneme and, to some extent, for the /z/. The first part is designed to shape the /s/ in isolation and then to incorporate it into consonant–vowel combinations and into words of sentences. In the second part the previous responses are strengthened, and the /s/ is incorporated into initial, medial, and final positions of polysyllabic words. Part three of the program is designed to strengthen the responses of the previous parts and to extend the production of /s/ into the connected speech of stories. A criterion test is administered at the termination of each of these subprograms to determine whether the client should go on to the next part or repeat the same items.

The American Book Company now publishes speech articulation kits for /s/, /l/, /r/, and /th/. This commercial development follows the Mowrer S-Pack format, with some modifications and improvements, including assessment and recording of progress from session to session. Similar articulation programs for ten phonemes have been developed for aides of speech clinicians in Allamakee, Howard, and Winneshiek counties of Iowa (Lubbert et al., 1973). Initial, carry-over, and home program phases are included for /s/, /z/, /θ/, /ð/, /f/, /ʃ/, /tʃ/, /k/, /r/, and /l/.

A simpler and yet effective program for transferring a target behavior, once it has been shaped in an original stimulus, has been developed by McLean (1967). The clinician presents a series of words containing the target phoneme under echoic, echoic–picture, picture, picture–word, word, word–sentence, and sentence conditions. The words are presented under each condition until a criterion is reached. When the final criterion is met, additional words are presented to test for generalization of response. Such programs for shaping and transferring target responses can be developed for any speech behavior.

Bokelmann (1973) and her associates have developed six instructional material manuals for the TH sound, the L sound, the S and Z sounds, the SH and ZH sounds, the CH and J sounds, and the R sound. Each program is designed to transfer correct production of a target phoneme from the therapy room to the principal's office in 15 steps. The client generally progresses from speech activities in the therapy room, to a quiet outside setting and hallway, to a classroom and a busy outside area, and finally to the school office and the office of the principal, who awards a certificate of program completion. Various personnel, including the clinician and an aide, interact at various parts of the program. Activities include naming pictures, making up sentences, recalling pictures, telling stories, and answering questions. Criteria are established for entrance into the program and for the 15 steps of the program. Clients exchange tangible reinforcers for items from a menu of rewarding activities. A scoring sheet records correct and incorrect responses for each step.

Electronic, computer, and operant technology have been utilized by Holbrook (1972) to shape and refine pitch, nasality, and intensity of speech. Digital displays, meter needles, lights, beeps, and events provide readouts and reinforcers. Baseline measures on clients are called habitual levels; final training measures, conditioned levels; posttraining measures, extinction levels; and measures taken 3 months after training are referred to as postexperimental levels. A "Charlie Brown" output attachment advances when a young child is progressing toward the criterion; this attachment kicks over a cup containing a reinforcer when the criterion is reached. With one device, if the child speaks too loudly, he hears an unpleasant beep. Appropriate reinforcement schedules are utilized to achieve and maintain a high rate of response within the criterion range. The procedure is repeated for more and more precise speech criteria.

Fletcher (1972) has designed similar instrumentation called TONAR II. A client can evaluate his own progress on a reinforcement panel that includes white, blue, and green lights. A white light is lit briefly when a spoken response reaches the specified criterion level. One of ten blue lights is lit when the number of successes reaches the number preset on a criterion switch. The ten lights are positioned diagonally up the panel. If the criterion setting is 2, for example, the blue lights move one space after the white light has blinked twice. When the tenth blue light is lit, one of a series of ten green lights, serving as decade counters, is triggered. The green lights are positioned along the bottom of the display panel. Together with the blue lights, the green lights inform the client of his total successes. Each time a succeeding green light is triggered, the next success is signaled by the first or lower left blue light. The control and display console of TONAR II also includes four counters: total trials, total successes, criterion trials, and criterion successes. Total reset and separate reset switches permit the clinician to track successes within different criterion settings.

Technical and marketing problems have interfered with the widespread incorporation of Holbrook (1972) and Fletcher (1972) instrumentation into clinical applications. When such sophisticated analyzers and programmers do find a common place in the management of speech, they may be even more useful for monitoring remediation than for directly facilitating it. Continuous or periodic monitoring of various parameters of speech can serve to track clinical interaction effectiveness and progress.

A precision program developed by Johnson (1974) to reduce vocal abuse is a model for monitoring progress during speech management. Johnson's Vocal Abuse Reduction Program (VARP) (1) pinpoints vocal abuse behaviors for each client in whom such behaviors have resulted in the formation of laryngeal pathologies, (2) systematically reduces these behaviors in specific high probability situations or time periods, and (3) reduces the size of or eliminates laryngeal pathology, making possible the establishment of normal vocal quality.

Lindsley (1971) describes features of precision teaching, which underpins the development of the VARP program. He credits Skinner (1953), the discoverer of operant conditioning, with the innovation of recording frequency of response, which is called rate. Lindsley states that every behavior has a frequency and that changes in frequency over time, called acceleration if positive or deceleration if negative, are universal and easily read from a behavior chart. The data for the chart are counted by the client, so there is no problem of observer reliability. The response that is being counted and graphed is the same behavior that is being manipulated and reinforced. Synthetic or extrinsic consequences do not need to be considered.

Diedrich (1973) has developed a program designed to teach a clinician how to utilize the Lindsley chart to record data on progress in speech remediation. Six steps are followed in counting and charting target behaviors: (1) define the critical behaviors to be charted; (2) define what is to be accepted as right and wrong; (3) decide the time base; (4) decide how often the behavior will be charted; (5) decide by what means the responses will be elicited; and (6) decide in what environment the sample will be taken.

Methodology in Educational Audiology

During the past decade, the term *educational audiology* has become prominent in the practice of the habilitative management of the hearing-impaired child. The term generally suggests increased utilization of residual hearing in the educational management of the hard-of-hearing youngster. Of all senses, audition offers the greatest potential for contributing sensory information about voice, prosody, and articulation for the great majority of hearing-impaired children in special classes and in regular classes.

In the development of educational audiology, Berg (1976) has drawn upon the newer scientific and instructional technologies to design speech management programming for the hearing-impaired client. He divides speech management into five stages: evaluation, sensory training, shaping and refinement, transfer, and generalization testing. A detailed description of many educational audiology contributions to speech management is presented in Chapter 6 of Berg's book *Educational Audiology: Hearing and Speech Management* (1976). Critical features of these contributions are described below.

Speech Evaluation

The evaluation of the speech of a hearing-impaired client provides a baseline for initiation of optimal training procedures. The objectives of speech evaluation include the identification and description of speech errors. Specifically, these errors may be omissions, substitutions, distortions, or additions of vocal, prosodic, and articulatory features of speech. If the hearing-impaired client is deaf, the clinician may anticipate errors in five general categories: timing and rhythm, pitch and intonation, nasality or denasality, articulation, and voice quality and loudness (Nickerson, 1975). When the client is hard-of-hearing, his speech errors characteristically are limited to hypernasality and misarticulation (Berg, 1970).

Ling (1976) has developed a comprehensive speech evaluation for a hearing-impaired client. This evaluation includes (1) an oral peripheral examination of the structure and functioning of the speech organs, (2) a preliminary phonological evaluation of a 50-utterance tape sample, and (3) a determination of whether many essential speech subskills are present in a client's phonetic repertoire. In the phonetic evaluation, a clinician indicates whether a series of target behaviors are produced consistently, inconsistently, or not at all. The phonetic evaluation encompasses vowels and diphthongs in isolation, simple consonants and word–initial blends in three syllables apiece, and word–final blends in syllables.

A unique SRAP marking system is also incorporated to varying extents in Ling's

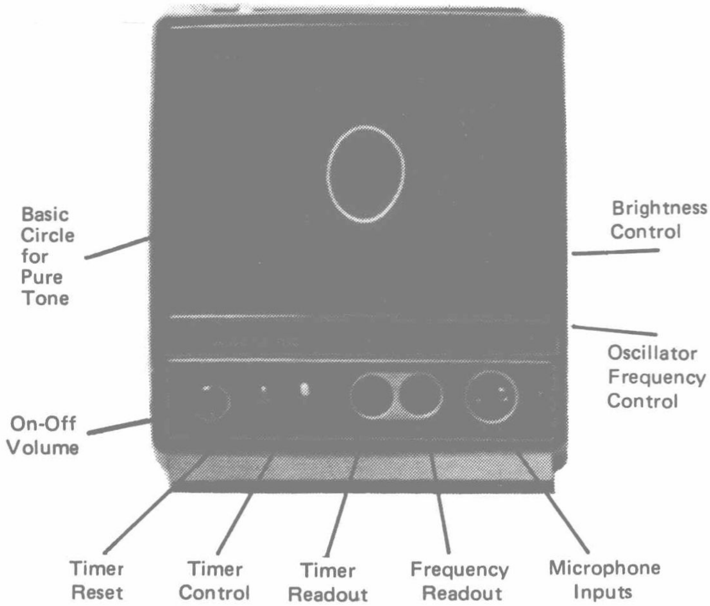

Figure 1 *Features of the Video Articulator*

phonetic evaluation. The *S* means that a client can sustain a single syllable containing the target behavior at loud, quiet, and whispered levels. The *R* is marked when the syllable can be repeated at a rate of three per second. *A* is checked when the same syllable can be alternated with any other syllable at the same rate. The *P* is checked when the client can vary pitch over eight semitones while producing the syllable containing the target behavior. Ling (1976) indicates that the absence of one or more of these subskills interferes with the feed-forward demands inherent in fluent speech.

The number of phonetic contexts included in various tests of articulation varies considerably. One commonly used word test targets on 12 vowels, 6 diphthongs, 25 single consonants, 64 double consonant blends, and 5 triple consonant blends (Templin and Darley, 1969). A Deep Test of Articulation by McDonald (1964) tests the articulation of 13 difficult-to-produce phonemes in as many as 49 abutting consonant contexts apiece. Abutting consonants, said quickly, may present an articulatory task comparable to producing consonant blends. The Goldman-Fristoe Test of Articulation (1972) includes 117 phonetic contexts in a sounds-in-words subtest and a sounds-in-sentences subtest. The phonetic evaluation of Ling (1976) includes vowels, diphthongs, simple consonants, 26 word–initial blends, and 52 word–final blends. The blends are sequenced into categories according to difficulty of articulation.

Berg (1976) has introduced a target format to assess the accuracy of a vocal, prosodic, or articulatory response. An omission, substitution, or gross distortion is marked off-target, a moderate or severe distortion, in the outer ring; a slight or mild distortion, in the inner ring; and a precise production, in the bull's-eye of the target. These degrees of accuracy are converted to numbers: 4 indicates an off-target response, 3 is for outer ring production, 2 equals an inner ring response, and 1 is equivalent to bull's-eye accuracy. An overall measure of accuracy can be computed by adding up the number of fours, threes, twos, and ones and dividing this sum by the number of speech responses being assessed. The assessment of accuracy of response differs from the usual dichotomous test procedure of marking an item either correct or incorrect. A dimensional marking procedure seems to be particularly appropriate for a hearing-impaired client because such speech often includes a great many distorted productions. These speech distortions reflect degrees of inaccuracy in a client's auditory discrimination of vocal, prosodic, or articulatory phenomena.

The assessment of accuracy of a prosodic speech response is also aided by the development of a refined system for marking the stress and intonational phenomena of speech. The stress and intonation notations developed by Berg (1976) permit the client to compare the detail of standard stimuli with prosodic responses of a client. If the notational patterns for the clinician's stimulus and the client's response do not correlate at all, the response is off-target. Patterns that correlate exactly reveal a client response that is precise or in the bull's-eye of a target. Patterns with varying degrees of correlation correspond to outer ring or inner ring accuracy of response. Recording stress and intonation markings prior to assigning target accuracy may increase the reliability of the assessment of response.

Once the accuracies of vocal, prosodic, and articulatory responses are recorded on targets, the clinician finds it valuable to display these data so the many errors can be viewed in comparative relationship with one another. A distinctive feature analysis is included as a component of this summary display. Berg (1976) has identified 11 distinctive features and corresponding definitions in his book *Educational Audiology: Hearing and Speech Management*. Most of these are included in the McReynolds and Engmann (1975) distinctive feature system being utilized in speech pathology. The additional distinctive features identified by Berg (1976) are *squaring* of the lip orifice and *buzzing* or vibration of the tongue or lower lip, each phenomenon

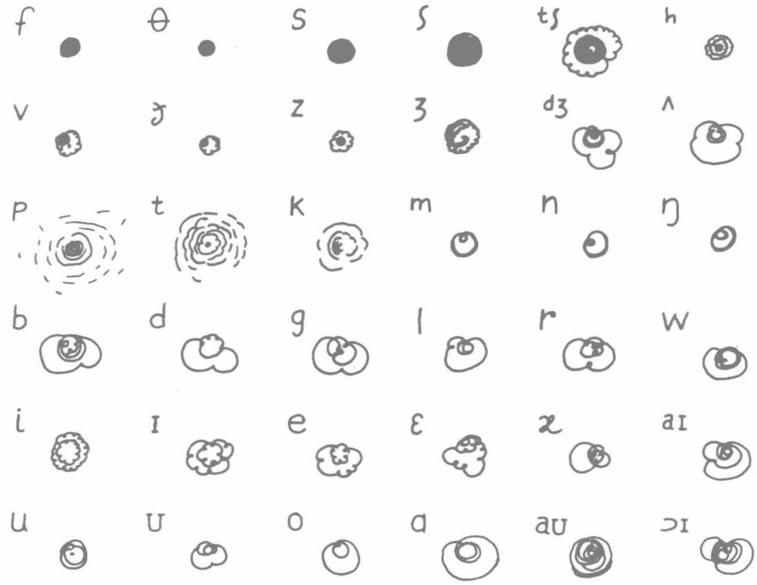

Figure 2 *Freehand drawings of videograms for 36 isolated phonemes*

occurring with articulation of specific phonemes.

Sensory Training

Following speech evaluation, and prior to shaping and refinement, Berg (1976) recommends that a clinician determine the extent to which a client utilizes speechreading, tactile, auditory, and electrovisual clues in the perception of a vocal, prosodic, or articulatory skill in error. Called sensory training, the procedure is to repeatedly present a stimulus for a target behavior, together with one or more other stimuli, requiring imitative responses, or pointing to corresponding pictures and recording correctness. Baseline and training scores are obtained under auditory, facial (speechreading and/or tactile), and electrovisual conditions. Ordinarily, a client's baseline performance in perception of speechreading, tactile, auditory, or electrovisual clues is below his potential sensory identification. For example, a hearing aid or an electrovisual speech indicator provided by the clinician might be ignored by clients until they have received corresponding sensory training.

In a case study experiment, Berg (1976) set out to determine which sensory aid conditions might be suitable while training a young "deaf" adult to speak. During each of seven 45-minute sessions, he presented seven sequences of 36 isolated phonemes to the client. Each sequence was presented under a different single or combined condition. Prior to the experiment the client was conditioned to utilize speechreading clues but not auditory or electrovisual clues. By the end of the fourth session, the client responded correctly to each of the phonemes under each of the conditions. During this sensory training experiment, the client also improved in accuracy of speech articulation, although he was instructed to merely discriminate the stimuli. The clinician assessed correctness of each discrimination by listening as the client tried to repeat each stimulus presented. This experiment is detailed in the book *Educational Audiology: Hearing and Speech Management.*

When clients cannot hear or cannot hear well, they have sensory gaps that may prevent the development of precise speech. These gaps cannot be fully compensated for by speechreading, kinesthetic, and tactile clues because they provide largely surface information on speech and insufficient deep and refined vocal tract phenomena normally supplied by hearing. Precise speech requires that clients perceive and produce detailed stress and intonation patterns, specific voice–voiceless and oral–nasal distinctions, and refined manner and place of articulation. Auditory clues from residual hearing can be amplified to assist hearing-impaired clients in perceiving many but not all of the needed vocal tract clues. When sound is converted into an electrovisual display, a visible sensory mode becomes available to clarify the detailed speech clues provided through residual hearing, speechreading, taction, and kinesthesis.

Video Articulator

An economical and yet sophisticated electrovisual speech-indicating device is the Video Articulator,[1] a successor to the Voice Visualizer (Pronovost, et al., 1967). The Video Articulator is a modified Sony television set weighing 9 pounds. The screen dimensions are 7 inches by 7 inches. Unique video patterns, called videograms, appear on the screen as the speaker says sounds, words, and sentences into a microphone. When a sentence is the stimulus, the overall vocal, prosodic, and phonetic content appears as a series of looplike patterns that vary in size and configuration. When utterance is slowed down, stress, pitch, voice–voiceless and oral–nasal, and manner and place of articulation features may be visually identified, particularly with sensory training. Controls and readouts of the Video Articulator are shown in Figure 1.

Sensory clues of the Video Articulator are of value also with speech clients who

[1]This device is marketed by Amera Incorporated, Box 627, Logan, Utah.

do not have a hearing loss. In these instances, video clues reinforce and clarify auditory clues, which can be fully perceived. They draw the attention of clients to specific vocal, prosodic, and articulatory features of speech. They are particularly powerful visual aids because they are precise visible correlates of acoustic and motor features of speech.

The author's freehand drawings of videograms for 36 isolated phonemes are shown in Figure 2. The drawings identify features that characterize specific consonants, vowels, and diphthongs. For example, relative intensities of the voiceless fricatives /θ/, /f/, /s/, and /ʃ/ appear as whirls of corresponding sizes and densities. Distinctive feature analysis and spectrographic information help the clinician anticipate what the patterns should look like. The client is also aided by speechreading clues in identifying consonant sounds that sound alike and that have correspondingly similar videograms, for example, /θ/ and /f/ or /ð/ and /v/.

The syllables of speech are also identified as different videograms. For example, /si/, /sɪ/, /se/, /sɛ/, /sæ/, /su/, /sʊ/, /so/, /sɑ/, /sʌ/, /sɝ/, /saɪ/, /saʊ/, and /sɔɪ/ may be discriminated electrovisually from one another. Ling (1976) suggests that perception of a feature of speech is less important than identification of its interaction with other aspects of speech. In these and many other syllables, the component parts and the combined features can be perceived electrovisually.

Ling (1976) states that visual displays do little to foster feed-forward control of speech. He also indicates that visual indications may help a client monitor a speech pattern when produced but cannot teach him how to make the pattern in the first place. These criticisms hold true for any sensory input system, including audition as used by persons with normal hearing. Feed-forward is aided when a client can reflexively produce the essential subskills of vocal, prosodic, and articulatory behavior. Electrovisual feedback contributes dramatically to the sensory pool needed for teaching the numerous motor subskills of speech. Clients can learn to identify videograms they produce, associate them with desirable and undesirable productions, and eliminate patterns that should not be included in their developing speech repertoires. Rather than providing less information than alternative forms of feedback, as Ling also states, electrovisual clues such as videograms picture the detail of the acoustic features of speech that normal-hearing people rely upon to develop precise speech.

A videogram for a consonant, vowel, or diphthong generally can be seen in isolated production and in words and sentences, particularly when speech is slowed down. The exceptions are the stop and affricative consonants, which have very short durations when produced in speech contexts. Clues from taction, speechreading, or residual hearing, if available, reveal the presence of stops and affricates, for example, /t/ and /tʃ/, in words and sentences. A good example of videogram visibility in context is that of the /s/ phoneme. It can be readily seen by the client in isolated production and in words such as *so, us, nice, listen,* or *Mississippi.*

Videogram applicability in the teaching of feed-forward subskills specified by Ling (1976) may be exemplified in the instance of the fricatives /s/ and /z/. Initially these are taught in isolation. Then they are taught following the vowels /u/, /a/, and /i/ in VC syllables. The client is then required to produce each of these syllables repeatedly. These fricatives are then held for at least 3 seconds in the same syllables. The /s/ and /z/ are then produced in the intervocalic position of contexts like /izi/ initially with each syllable equally stressed and thereafter with one vowel stressed and the other not. The /s/ and /z/ are then produced in the initial position of single syllables like /so/ and /zu/. Then /s/ and /z/ are taught in such syllables as /pasapasa/ and /wiziwizi/. Finally, these fricatives are produced in final, medial, or initial position of syllables using an eight-semitone pitch variation. In all nine of these subskills, distinct videograms for /s/ and /z/ are readily seen if the client articulates them.

Ling (1976) suggests that priority be given to teaching a client to produce the pattern of a syllable or sequence of them rather than a specific feature of a speech sound. The motor and acoustic pattern of contextual speech includes a complex of overlapping vocal, prosodic, and articulatory behaviors. For the hearing-impaired client, none of the sensory input systems by themselves or in combination can provide the detail of this behavioral complex. A speech management program can initially assist the client to approximate a contexual speech stimulus for communicative applications. Thereafter, subskills can be taught so that syllables or series of syllables can be recombined more precisely. During initial shaping or later refinement, sensory clues from each of the input systems are utilized. No one sensory input system seems to be inherently any more or less distractive or misleading than any other one. Whether residual hearing, speechreading, tactile, kinesthetic, or video clues contribute or distract depends upon the design and implementation of the speech program being utilized by the clinician.

Berg (1976) has developed a terminology to draw attention to the detail of each videogram. Terms such as *circles, lobes, loops, whirls,* and *splashes,* as well as *concentration, tilt, compression,* and *rotation,* are helpful descriptors. If photographed, the videograms for the various consonants, vowels, and diphthongs look as much alike or different as their correlative sounds are alike or different. For example, the /i/ and /o/ look very different, and the /m/ and /n/ look highly similar. In addition, two persons with different voices will produce somewhat different videograms unless they are articulating voiceless consonants.

The Video Articulator may reveal subtle changes in the production of any phoneme or any vocal or prosodic parameter. Each /ɝ/ allophone articulated, for example, may appear as a noticeably different videogram. Acoustically noticeable pitch, loudness, durational, laryngeal, and velopharyngeal changes may also be seen.

Both microphone distance and vocal intensity determine the size of the video patterns appearing on the screen. If the speaker is too intense, the video patterns will be too large and extend beyond the boundaries of the screen. Both the clinician and the client need to practice microphone technique and use of precise speech input. When speech is carefully produced, videograms may be easily recognized. When the speaker is careless or imprecise, the video patterns correspondingly reflect gross and unpracticed input.

The Video Articulator has extra capability for pitch measurement and feedback. A FREQ (frequency) control can be turned on and the frequency of an oscillator varied from below 100 Hz to above 500 Hz. The oscillator frequency can be heard through the loudspeaker, seen as a single circle on the screen, and seen also as a digital readout of a counter. The video pattern is stabilized when the fundamental frequency (pitch) of the speaker's voice is tuned to the oscillator frequency. This "tuning" phenomenon may be recognized when a speaker utters

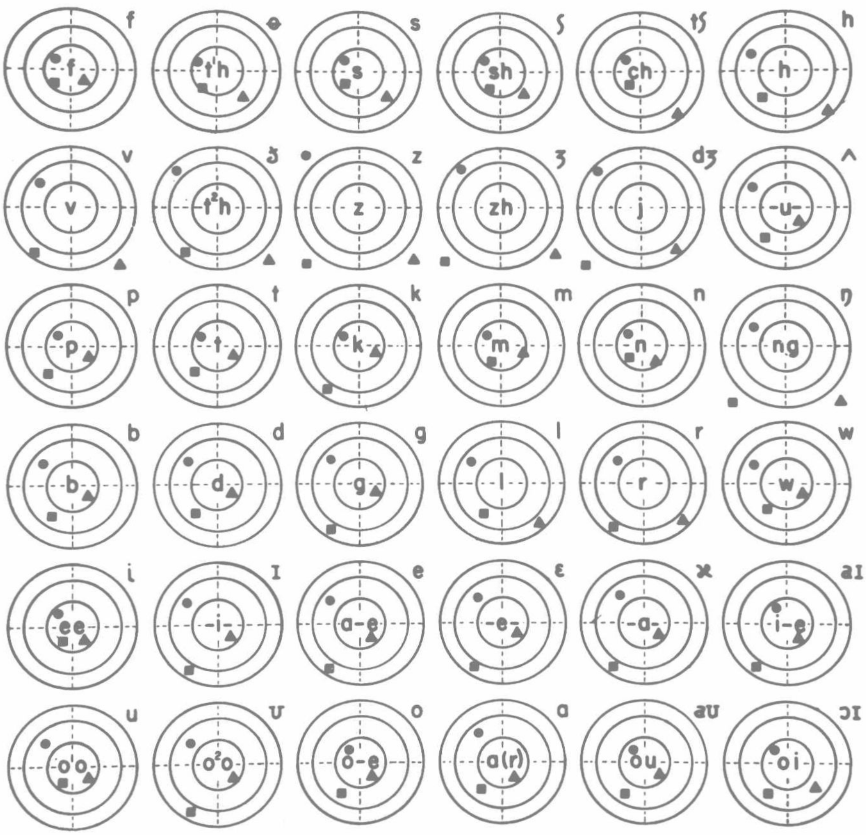

Figure 3 *Baseline accuracies of three hearing-impaired children in the articulation of 36 critical articulations*

a sustained sound or monotone connected speech. The FREQ control can be adjusted to locate one pitch or to bracket the entire pitch range of a client.

A digital stopwatch is also built into the Video Articulator to record durations of vocalizations and time for completion of clinical tests or tasks. The readout progresses from 1 second to 10 minutes and then repeats itself. When sustaining a phoneme, clients can watch video patterns they produce and the readout of progressing time. These times can be recorded to chart progress.

Shaping and Refinement

The third stage of a speech management program is the shaping and refinement of a vocal, prosodic, or articulatory behavior in error. Shaping and refinement consists of steps leading through mere target responding to consistent bull's-eye production. It precedes transfer of a response into other speech contexts, stimulus conditions, and real-life settings. In the shaping and refinement process, the clinician designs and manages antecedent and subsequent events. Before a client's response, the clinician provides instruction, demonstration, sensory clues, and stimulation. Immediately after the response, the clinician provides the client with reinforcement and feedback about correctness. Shaping and refinement of a behavior in error may be programmed within one stimulus or within a series of stimuli requiring that specific behavior. The stimuli may be behaviors in isolation, nonsense syllables, monosyllabic or polysyllabic words, phrases, sentences, or combinations of sentences such as occur in conversation.

The target format developed by Berg (1976) is an aid in the shaping and refinement processes as well as in speech evaluation. In an experimental program for three young hearing-impaired children, speech training time was spent on each of 36 phonemes during a given session. The baseline accuracies are shown in Figure 3. If a phoneme was already at bull's-eye accuracy, it demonstrated precision production, contributed to the overall success rate, and often included a distinctive feature that needed to be incorporated in one or more phonemes in error. A single target was divided into 36 pie-shaped sectors to record training data. The phonetic symbols for the 36 critical articulations were affixed to the sectors.

While training a client, a clinician modeled a given misarticulated phoneme, using a facial–auditory–vide (electrovisual) condition. The client imitated the clinician, and the clinician recorded a dot on or off the target to indicate the response's degree of correctness. The clinician also provided token and social reinforcement to reward improvements from off-target (4) to the outer ring (3), to the inner ring (2), and into the bull's-eye (1). Numerous repetitions of this procedure were presented. Before each stimulus presentation, the clinician pointed to the mouth, throat, nose, ear, or video screen to draw the attention of the client to an appropriate clue. All the client's responses were recorded as dots. The best response for each phoneme provided the basis for affixing a score of 4, 3, 2, or 1. The mean baseline accuracy for one client was 2.8. The corresponding

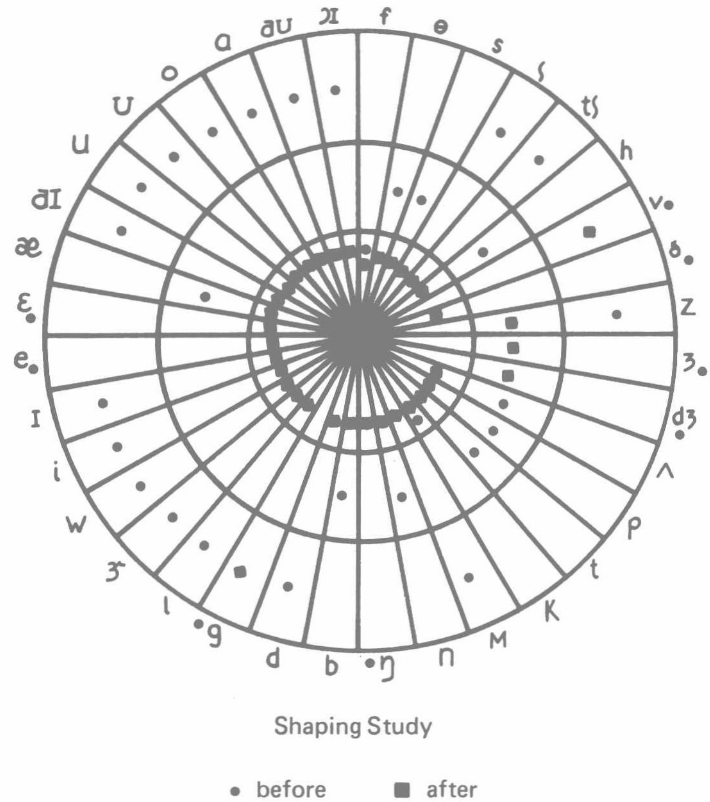

Figure 4 *Baseline and final measures of accuracy in the articulation of 36 isolated phonemes by a young hearing-impaired child*

posttraining measure after ten blocks or presentations of 36 phonemes was 1.2. The baseline and final measures of accuracy are shown in Figure 4. Preliminary sensory training was given to condition the client to utilize video clues in addition to other sensory input.

Various target options may be designed by a clinician to record shaping and refinement data. A single target can be divided into small sectors, as seen in Figure 4. Another option is use of a separate target for each possible speech error. Data on each speech response can also be recorded on a large target and transferred as a single datum to one of the previous types of targets. Data on evaluation and sensory training can be kept on the same targets where shaping and refinement data are recorded.

A clinician will find that his vocal, prosodic, or articulatory stimulus can be a useful model during initial shaping or when the goal is to move a production onto a target. One or more of kinesthetic, tactile, speechreading, auditory, and electrovisual clues accompany each stimulus. The clinician utilizes an appropriate combination of these clues to prompt clients to achieve and stabilize a target response. Afterward, clients' random or prompted improvements become the standard they are trying to reproduce. The clinician's task, then, is to listen critically and indicate on a target how closely clients approach their desired productions. In time, clients learn to recognize and monitor their own hits and improvements. The combination of various sensory clues are helpful in shaping an error behavior onto the target. The auditory and particularly the video clues may be especially helpful, together with target feedback, for prompting the client to inner ring and finally bulls'-eye precision in speech production.

The role of tactile-kinesthetic (TK) feedback for shaping speech behavior may be better understood from two experiments conducted by Ling and Bennett (1974–1975). In each experiment, two young, profoundly hearing-impaired children were subjected to vowel training through operant conditioning procedures. In the first experiment the children were trained to imitate the isolated vowels /a/, /æ/, /o/, /i/, and /u/ using an audiovisual (AV) (V = speechreading) feedback. The number of trials required were 75, 150, 50, 500, and 25 and 75, 25, 50, 225, and 25, respectively. The /i/ was the most difficult to learn. The audiometric evidence suggested that neither child seemed to hear the second formant for the /i/.

In the second experiment, two other children were trained to produce the /i/ and /au/ combined with the initial /p/ consonant. With the /pi/, early training utilized TK feedback and with the /pau/, AV feedback. In the TK condition, the child placed a finger of one hand in his mouth and the corresponding finger of the other hand in the experimenter's mouth. These fingers were placed on the tongue to feel position or placement. The data revealed that initial TK training of these children facilitated progress in learning to imitate high front and high back tongue placement.

It would be interesting to replicate these experiments, using children with varying degrees and slopes of severe and profound hearing impairment, to assess the value and permanence of initial conditioning with TK stimulation during the shaping and refinement of consonants, vowels, and diphthongs. Such stimulation may be particularly applicable to guiding a phoneme onto target, whereas AV conditioning may be especially helpful in refining accuracy, and electrovisual feedback may aid in facilitating bull's-eye production.

The ease and speed with which a consonant, vowel, or diphthong, or a vocal or prosodic behavior, is shaped onto a target and refined into a bull's-eye vary from sound to sound and from client to client. For example, the /p/ and /t/ and the /f/, /θ/, and /h/ can be shaped and refined relatively quickly, notwithstanding hearing loss. The facial (speechreading and tactile) clues are sufficient and the motor act relatively simple for these voiceless articulations. Auditory clues and particularly videograms, when used with target feedback, assist in further refinements and stabilization of precise response. Other target behaviors are more difficult to perceive and /or to produce and require more training. Motivational, perceptual, and conceptual differences among clients also influence ease and speed of learning.

The clinician should shape and refine vocal, prosodic, and articulatory behavior within the context of an overall speech development or corrective program. The Ling (1976) training model provides a structure that can be utilized for a client regardless of age, hearing impairment, or current speech competency. Globbal exposure to speech may enable the client to acquire successive target behaviors without specific training with subskills. Periodic evaluation enables the clinician to determine when intervention at a subskill level needs to be undertaken. The curriculum does not require that specific targets be taught by particular ages or grades. Speech acquisition is considered globally as a series of phonetic and prosodic stages, a group of carefully defined target behaviors, some of which are trained concurrently, and many sets of criterion-based subskills.

A multisector target may be used to record incremental progress in the management of any series of speech behaviors. Berg's (1976) critical articulations or one of Ling's subskills series may be the target behaviors. Figure 5 provides a 36-sector target that can be used for these and many other applications. Initially, the clinician labels each sector and then duplicates copies. One copy may be used to record baseline data. Additional copies record dots for successive training blocks. The baseline and final training data can be compared on the last copy of the target. The final training data may demonstrate meeting the terminal pass criteria. The clinician can save paper by recording successive training blocks with different colored dots on a single copy. A key can be written on the target to correlate colors and training blocks.

An option to tracking a series of behaviors on one target is using a target for each behavior. The 36 targets of Figure 6 exemplify this format. Each target is divided into 12 pie-shaped sectors. Baseline data are recorded in the 12 o'clock to 1 o'clock sectors, and block-by-block training data are recorded in the succeeding sectors of each target. The pattern of data from sector to sector reveals training progress.

The clinician must determine criteria for bull's-eye, inner ring, outer ring, and off-target accuracy when using the targets in either Figure 5 or Figure 6. Ordinarily, bull's-eye means a precisely correct response. Off-target indicates a response so incorrect that no correlation exists between stimulus and response. Inner ring and outer ring accuracies fall correspondingly between bull's-eye and off-target.

Transfer of Training

Once a vocal, prosodic, or articulatory target behavior has been shaped and refined in an original stimulus or in a series of subskills, a need exists to transfer this skill into additional stimuli and stimulus conditions that constitute or simulate real-life experiences. Such carry-over programming is indispensable to effective interpersonal communication and should be implemented by the educator of the hearing impaired or by the speech pathologist. Without transfer training, the client finds overall speech work to be less meaningful and functional.

Three approaches to transfer programming are described in the literature. One is the oral approach of education of the hearing impaired that provides correction of speech outside of remedial sessions (Calvert and Silverman, 1975). The second approach is exemplified by programming kits or packages utilized by many speech pathologists (Bokelmann, 1973; Lubbert et al., 1973; Mowrer et al., 1970). The third approach to transfer programming is the McLean format, which trains the target behavior in up to ten additional stimuli and in a series of simulated conditions. In the oral approach, transfer of training is facilitated considerably. In the kit method, syllables and many words and sentences are used to transfer the target behavior. The McLean format includes fewer stimuli but incorporates them repeatedly.

The McLean format may be the most feasible approach for systematic transfer of a large number of vocal, prosodic, and articulatory behaviors. The stimuli and stimulus conditions for any behavior or series of related behaviors can be packaged in relatively little space. Therefore, the stimuli and recording forms for many needed transfer applications are readily available to the clinician at a minimal cost. These transfer programs can be criterion-referenced to facilitate learning and enhance accountability.

Generalization Testing

During and following a speech management program for a specific error, a clinician may test for generalization of the trained target behavior. The sound production tasks of Shelton and his associates (Elbert et al., 1967; Shelton et al., 1967; Wright et al., 1969), a follow-up to McLean programming (White, 1972), and the Bokelmann (1973) program provide three models for relevant generalization testing.

Each Shelton sound production task includes 30 or more stimuli that incorporate a specific target behavior. The /s/, for example, is included in three different orders of 30 items. The stimuli incorporate the /s/ in isolation, in nonsense syllables, in monosyllabic words, in polysyllabic words, and in phrases and sentences. The items are presented during successive training sessions to trace progress from the beginning of an /s/ training program to the end (Shelton et al., 1967).

Generalization tests following a McLean program may be of three types: new item, across position, and overgeneralization. A new-item generalization test will incorporate additional stimuli with the target behavior in the transfer program position of words. The across-position generalization test will include words with the target behavior at the other end of stimulus words. An overgeneralization test is only used when a substitution error has been involved. During transfer training, the substitution has