Chapter 13

Current and Future Directions in Research on Speech Intelligibility Assessment of Persons who are Deaf

Dale Evan Metz and Nicholas Schiavetti
State University of New York – Geneseo

This chapter considers some major issues in the assessment of speech intelligibility deficits exhibited by persons who are deaf. Major themes concern the relative importance of factors affecting speech intelligibility, the relationship of physical speech parameters to perceptual speech intelligibility measures, and the reliability and validity of various approaches to the measurement of speech intelligibility. Suggestions concerning future research directions are predicated upon the implications of current research for computer-based measurement of both physical speech parameters and their ultimate use for speech intelligibility estimation. The research findings discussed in this chapter indicate the importance of careful attention to methodological issues in the assessment of speech intelligibility.

The role of speech in the communication of persons who are deaf has long been a controversial issue (Davis & Silverman, 1978). Prior to the development of the total communication philosophy, many schools for deaf children were characterized as exclusively "oral" or "manual" in their approaches to the education of deaf students. Within the total communication framework, schools paid varying attention to both oral and sign language communication needs, preferences, and abilities of deaf students. In recent years, the controversy regarding communication modes used by deaf persons has revived concern regarding the preservation of Deaf culture;[1] most notably communication in American Sign Language or ASL (Padden & Humphries, 1988).

Within the context of this communication controversy, it must be acknowledged that some deaf persons will choose to communicate primarily via speech, some primarily via sign language, and some will choose varying degrees of both speech and sign language communication. An important viewpoint that has emerged from Deaf culture argues that the decision to communicate by speech (and, thus, the contingent decision to seek speech improvement services) must rest with the individual deaf person. The purpose of this chapter is not to arbitrate conflicting views about preferred modes of communication among deaf and hearing persons. Rather, the purpose of this chapter is to review some recent research findings and directions regarding the relationship between selected acoustic parameters and overall speech intelligibility of deaf persons. The research discussed in this chapter has, in large measure, contributed to our current efforts regarding the development of a robust test of speech intelligibility that is predicated on the significant underlying dimensions of intelligibility deficits.

Emphasis is placed on the term "explanatory speech intelligibility testing," a concept alluded to by Monsen (1978) regarding the speech of deaf persons and formalized by Weismer, Kent, Hodge, and Martin (1988) and Kent, Weismer, Kent, and Rosenbek (1989) regarding the speech of dysarthric persons. This term refers to speech intelligibility measures that are based on linguistic and acoustic parameters that could potentially serve to explain the intelligibility deficit in articulatory terms. We will argue that speech intelligibility is uniquely determined by an individual's ability to accurately produce a *finite set* of linguistically significant acoustic contrasts that relate to segmental and suprasegmental distinctions. These critical distinctions are frequently not captured adequately by extant assessment techniques that rely on single or multiple productions of isolated segments and syllables. Evidence suggests that some speech production assessment techniques fail to achieve distinctiveness among segmental and suprasegmental subtests when mathematically scrutinized.

A central theme throughout this paper is that we are attempting to understand the physical, psychophysical, and perceptual components of speech intelligibility. The complete delineation of these components of speech intelligibility will improve understanding of the speech production skills of persons who are deaf. As Kent (1992) has stated: "Intelligibility is the *sine qua non* of spoken language. Disorders that impair intelligibility are among the most serious disorders of communication" (p. 9).

---

[1]Throughout this chapter we employ the convention proposed by Woodward (1972). We use lowercase "d" in deaf when referring to an audiological condition and uppercase "D" in Deaf when referring to Deaf people who share a language (ASL) and culture (cf. Padden & Humphries, 1988).

In this regard, we seek a level of generality in our descriptions of speech production characteristics and assessment protocols that transcend and subsume surface level categorizations of speech error types. We have previously argued (Metz, Samar, Schiavetti, Sitler, & Whitehead, 1985; Metz, Schiavetti, Samar, & Sitler, 1990) that understanding speech production characteristics of persons who are deaf (and hence assessment and training) would be well served by discovering the integral character and unique origins of speech production dimensions that veridically relate to an individual's speech intelligibility. It is our belief that common origins are more readily discovered, understood, and manipulated than multivariate origins.

In any discussion of the speech communication of deaf persons primary consideration must be given to the specific speech characteristics that will be assessed and the relationship of assessment techniques to the outcome and efficacy of training programs. Our intention is to consider mainly the perceptual criterion of speech intelligibility and the relationship of this perceptual criterion to specific underlying acoustical parameters of speech production. The rationale for this decision is principled by virtue of the fact that speech intelligibility is the single most important criterion to apply in analyzing both the success of speech instruction outcome and the relative efficacy of instruction programs.

Subtelny (1977), for example, stressed the importance of intelligibility measures when she stated, "Intelligibility is considered the most practical index to apply in assessing competence in oral communication" (p. 183). Additionally, Monsen (1981) has suggested several important uses of speech intelligibility assessments with deaf children, including monitoring progress in speech instruction, comparing methods of speech training, and evaluating candidates for mainstreaming. Intelligibility assessments also play a role in pre- and postoperative evaluations of persons with cochlear implants, although such assessments apparently are not relied upon as heavily as other forms of speech production testing (cf. Beiter, Staller, & Dowell, 1991; Tobey & Hasenstab, 1991). This state of affairs is arguably related to the fact that current methods for intelligibility assessment (despite the unequivocal importance of global intelligibility measures) fail to achieve a suitable level of generality and clinical utility beyond a common index of relative degree of speech dysfunction. As Weismer et al. (1988) point out, intelligibility measures typically do not contribute to one's understanding of the basis of the intelligibility deficit beyond the general indexing of severity levels. Clearly, potentially subtle changes in speech production skill associated with persons with new cochlear implants could go undocumented by such broad classification schemes. We will offer some preliminary developments of a speech intelligibility test that will go well beyond such indexing and, perhaps, obviate the use of certain extant speech production tests that have questionable psychometric properties.

THE SPEECH INTELLIGIBILITY CONSTRUCT

What is speech intelligibility? We are in general agreement with Carney's

(1986) definition that states "the term [speech intelligibility] may be defined generically as that aspect of oral speech-language output that allows a listener to understand what a speaker is saying" (p. 47). Implicit in this definition is that not all aspects of vocal behavior during speech are relevant to linguistic communication (i.e., speech intelligibility). Accordingly, speech intelligibility measurements should be designed to reduce or eliminate the contaminating influence of the irrelevant acoustic properties of speech on the perceptions (ratings) of listeners. (For an excellent discussion regarding variables that affect speech intelligibility ratings see Boothroyd, 1985.) Equally important is the notion that any assessment of overall intelligibility should be designed to provide recourse to the explicit acoustic dimensions that underlie specific intelligibility deficits. That is, the test should be designed to have an explanatory function (cf. Weismer et al., 1988).

Nickerson and Stevens (1980) recognized this need over a decade ago when they suggested that the measurement of speech intelligibility should

> . . . determine what there is, not only about the speech of a particular individual, but about a particular utterance, that makes it unintelligible, and to specify exactly how its physical characteristics would have to be changed in order for its intelligibility to be assured. (p. 338)

Addressing the concerns raised by Nickerson and Stevens (1980) requires the convergence of two sets of descriptive data: one set describing listener assessed speech intelligibility and a second set describing the speech errors of persons who are deaf. We strive to understand speech errors *at the acoustic level of measurement* because a small set of acoustic cues signal linguistic distinctions in English. These acoustic parameters could be useful in describing the extent of error and of changes that occur during the therapeutic process. The relationship of the above two data sets by mathematical equation forms the basis of an intelligibility test with explanatory power. The next two sections of this chapter elaborate the two data sets. Specifically we will examine (a) some of the procedures employed to assess speech intelligibility of deaf persons, and (b) the relationship of selected acoustic parameters associated with disordered speech to speech intelligibility.

SPEECH ASSESSMENT TECHNIQUES

The first step in the development of an explanatory test of speech intelligibility is to select or develop a reliable, valid, efficient, and flexible procedure to obtain information regarding speech production skills. In general, there are four popular assessment techniques that are currently employed to analyze speech production performance of persons who are deaf: (a) speech production skill rating scales, (b) segmental level speech production assessments, (c) forced-choice paradigms, and (d) sentence transcription procedures. All four of these procedures could be used to provide information for the first data set. But, the empirical evidence discussed below virtually disallows, on psychometric and psychophysical grounds, the first two forms of testing whereas it strongly suggests, on psychometric and psychophysical grounds, the tractibility of the latter two forms of testing.

Rating Scales

Subtelny (1977) developed six discrete rating scales designed to assess independently six dimensions of speech production: speech intelligibility, pitch control, pitch register, speaking rate, breath control, and prosody. In an effort to determine the efficacy and validity of this multi-level speech production test, Samar and Metz (1991) reanalyzed a large set of correlational data published by Subtelny (1977) utilizing the mathematical factor analytic technique of Principal Components Analysis (PCA). These correlational data reflected the relationships among the six independent tests. If these rating scales of segmental and suprasegmental aspects of speech production were valid and specific assessment tools, then their factor structure should reflect this multidimensional nature (i.e., six independent factors should emerge from the PCA if the tests are truly independent instruments). The underlying factor structure of six speech rating scales that emerged from the PCA was unidimensional. The dominant factor influencing all the measures in this analysis was articulatory integrity, despite the assumed focus of the rating scales on such factors as pitch control, pitch register, speaking rate, etc. These results suggest that raters may be incapable of suppressing competing perceptual information when trying to assess specific dimensions of complexly disordered speech.

Rating procedures are also used widely to assess the speech intelligibility of persons who are deaf, and are frequently the instrument of choice by virtue of their putative reliability, validity, and time economy. These presumptions may be misguided. A case in point regards a popular interval rating procedure for assessing speech intelligibility developed at the National Technical Institute for the Deaf (NTID) in the 1970s.

The NTID Speech Intelligibility Rating Scale, developed by Subtelny (1977), has enjoyed widespread use in both clinical and research applications (cf. Kelly, Dancer, & Bradley, 1986). Interval rating procedures are, however, only appropriate for psychophysical dimensions that exhibit substitutive properties along their continuum, like pitch perception (Stevens, 1975). Psychophysical dimensions that exhibit additive properties along their continuum, like loudness perception, are inappropriate for rating procedures. The exhaustive research conducted by S.S. Stevens and his colleagues (Stevens, 1975) has demonstrated that most psychophysical continua are additive, thus raising the question regarding the psychophysical dimension of speech intelligibility; is it substitutive or additive?

Schiavetti, Metz, and Sitler (1981) examined the psychophysical dimension of speech intelligibility utilizing the methodological protocols developed by Stevens (1975). The results of this examination revealed that the speech intelligibility continuum was classically additive in nature. There was a clear demonstration of the tendency for judges to subdivide the lower portion of the intelligibility continuum into smaller intervals than those at the upper end, suggesting nonlinearities in the scale. These findings indicate that the construct validity of the NTID rating scale was compromised. Additional evidence that directly demonstrated nonlinearities in the scale was provided by Samar and Metz (1990). When compared to well established transcription procedures, the NTID scale revealed substantial violations of measurement prediction within the clinically most frequent mid-range of speech intelligibility.

Interestingly, Samar and Metz (1990) reported very high intra- and interjudge reliability coefficients for the NTID rating scale, .97 and .94 respectively. These high reliability coefficients are, however, misleading when one considers the underlying mathematics. In engineering terms, many rating procedures used for assessing speech intelligibility (e.g., one to five or one to seven equal appearing interval scales, etc.) are categorized as having a large "least count." Least count is the unit of measure of the scale. It can be convincingly demonstrated mathematically that as least count increases, precision of measurement is lost, information about the precision of the measure is lost, and the ability to gain additional precision by averaging multiple measurements is also lost (Woelfel & Fink, 1980). The loss of precision is further compounded by the potentially deluding reliability estimates obtained by repeated measures. Large least count scales produce numerically smaller standard deviations giving rise to the impression of high measurement reliability. This occurs, quite simply, because, although random error is not eliminated by large least counts, evidence of such error is eliminated. Thus, large least count scales usually yield the high numerical estimates of proportional reliability (Woelfel & Fink, 1980). As Woelfel and Fink (1980) succinctly state:

> One can always increase the ratio of reliable to unreliable variance by the simple expedient of reducing the total amount of variance detected by the scale. This is misleading, since, other factors equal, that scale is most precise which produces the most variance or discrimination among the objects scaled. (p. 92)

The evidence demonstrating the difficulties with construct validity and potentially inflated reliability estimates strongly suggests that rating procedures should be avoided when one is attempting to accurately assess the speech intelligibility of deaf persons.

Segmental Production Assessments

Tests that purportedly measure specific segmental production skills are generally regarded to be reliable and valid instruments and the scores derived from such instruments have a direct relationship to a speaker's intelligibility. One of the most frequently employed instruments for such assessment with deaf children is Ling's (1976) Phonetic Level Evaluation (PLE). Recent evidence, however, suggests a lack of ecological validity of specific surface level ratings. Tye-Murray and Kirk (in press) compared the performance of deaf children on imitated vowel productions from the PLE and their vowel productions during spontaneous speech productions. Although many of the correlations between the PLE and spontaneous speech productions were significantly different from zero, the actual bivariate coefficients were low to moderate suggesting that a considerable amount of the variance associated with spontaneous speech productions was unaccounted for by the PLE.

Low correlation between the PLE and spontaneous speech intelligibility is further confounded by low inter-rater reliabilities. Very low intraclass correlation coefficients (ICC) on the PLE have recently been reported by Shaw and Coggins (1991), with only one of the four subtests, Simple consonants (Step 2), exceeding an ICC value above .80. Shaw and Coggins (1991) offer several plausible explanations for these low coefficients that do not appear to be easily redressed and rectified.

Forced-Choice Paradigms

Boothroyd (1985) presents an elegant demonstration that forced-choice testing protocols can be designed to virtually eliminate many of the unwanted mitigating influences on intelligibility estimates in addition to providing analytic details regarding segmental production characteristics. Such attributes are clearly essential for the development of an explanatory test. Moreover, intelligibility estimates derived from his forced-choice procedure are "reasonably predictive of the intelligibility of speech produced in more natural communicative settings" (Boothroyd, 1985, p. 185). Specifically, forced-choice test performance accounted for 70% of the variance in scores derived from structured sentence transcriptions. An important consideration when considering forced-choice testing is that the paradigm is particularly well suited for testing young deaf children as demonstrated by Monsen (1981). In addition, Boothroyd (1985) demonstrated that forced-choice testing greatly reduced the disparity between performance of inexperienced versus experienced listeners in intelligibility estimations of speech produced by deaf persons.

Despite the clear theoretical tractibility and clinical utility of forced-choice testing, properly constructed sentence transcription procedures appear to hold considerable validity as a method for speech intelligibility estimation. The empirical rationale for this assertion is presented below.

Sentence Transcription Procedures

For estimating speech intelligibility, transcription procedures appear to have significant advantages over rating and segmental analysis procedures, and a slight advantage over forced-choice paradigms. One major advantage is that sentence transcription procedures use a frequency count of the number of words on the listener's response list that match the words on the speaker's list of intended words. This frequency count is the intelligibility metric expressed by a percentage value or proportion of matched words. Thus, word identification tests produce a metric of speech intelligibility that is more readily usable by the researcher or clinician and percentages or proportions are easily interpreted by other professionals and laypersons. Indexing speech intelligibility with a percentage makes a certain degree of intuitive sense for informing someone of the degree to which they can expect to understand the speech produced by deaf persons. Additionally, word identification tests can be structured in sophisticated and revealing ways to permit precise assessments of far more than just articulation errors and the transcriptions themselves need not be limited to conventional orthography (Samar & Metz, 1991). It is in this regard that word identification tests appear to have their most significant advantage over other forms of speech production skill assessment.

Although an exhaustive speech intelligibility construct for deaf individuals has not yet been factorially delineated and validated, transcription paradigms have clear face validity with respect to the development of an explanatory test of speech intelligibility. For example, transcription procedures are selectively sensitive to the structural characteristics of speech-cue acoustic dimensions within the speech of the deaf. Additionally, transcription procedures are insensitive to normal variations in the vocal qualities of speech and to those abnormal variations like breathiness, abnormal pitch register, and the like that are perceptible but do not degrade the message intelligibility of speech. This specific selectivity is the direct result of the procedural constraint of forcing listeners to report exclusively their linguistic perceptual experience independent of their linguistically irrelevant perceptions of speech conspicuousness. Listeners are permitted to transcribe only the speakers' phonemic, lexical, or sentential utterances. In this regard, linguistic transcription procedures guarantee that such transcriptions will be based on the perception of acoustic cues of speech and not on other extraneous aspects of vocal productions (Samar & Metz, 1990). The relatively direct measure of linguistic experience afforded by transcription procedures is consistent with the general validation procedure asserted by Young (1969): "A measurement of a speech disorder is primarily a perceptual event, and the observer's response necessarily represents the final validation for any measurements" (p. 135).

Properly constructed transcription procedures also appear to possess superior intra- and inter-rater judgment reliabilities. In a large scale study, Samar and Metz (1990) asked 10 trained listeners to independently listen to 100 tape recordings of deaf persons' readings of the CID Everyday Sentences (10 subject tapes per listener). Each of the 10 speakers was assigned to each listener in such a way that each speaker audited by that listener had read a different one of the 10 CID Everyday Sentence lists. To assess listener reliability, approximately 2 weeks after the initial listening sessions, a subset of six tapes (three of which were previously audited by that listener and three of which had been audited by a different listener) were reassigned to the listening panel. Comparisons of the mean differences between the initial and replicate scorings revealed no significant intra- or inter-rater differences. The intra- and inter-rater reliability coefficients were estimated by computing the Pearson product-moment coefficients between the arcsin-transformed values of the initial and replicate scorings of each subsample. Both reliability coefficients exceed .98.

As indicated above, we are looking beyond the issue of basic overall intelligibility indexing to an intelligibility assessment that has explanatory power. The transcription paradigm offers the potential for such refinement. Transcription procedures can be internally structured so as to contain several factorially distinct dimensions of speech stimulus contrasts at different hierarchical levels of the structural control of speech and language (e.g., segmental vs. suprasegmental aspects). The transcriptional response constraints guarantee the extraction of these internal linguistic dimensions. Once done, one can extract the acoustic parameters underlying these internal linguistic dimensions to achieve a suitable level of generality for explanatory intelligibility testing. This brings us to the second data set that needs to be considered, description of the speech errors typically associated with the speech of deaf persons and their underlying acoustic representations.

ACOUSTIC CORRELATES OF SPEECH INTELLIGIBILITY

Objective speech measurements (acoustic and perceptual) have established that the speech of persons who are deaf is characterized by a broad variation in both the control of segmental and suprasegmental speech properties (Hudgins & Numbers, 1942; Monsen, 1978; Parkhurst & Levitt, 1978; Smith, 1975), and that these properties make *factorially* distinct contributions to the determination of overall intelligibility (Metz et al., 1985; Metz et al., 1990). Additionally, research using multiple regression analyses has demonstrated a strong predictive relationship between frequency counts of different types of surface level speech errors produced by deaf persons and overall ratings of speech intelligibility (Stromberg & Levitt, 1979). Monsen (1978) examined the relationship between selected acoustic speech parameters and speech intelligibility. The author reported a good predictive relationship (73% of the variance in overall intelligibility scores) between the acoustic parameters of the voice-onset time (VOT) differences between /t/ and /d/, the second formant difference between /i/ and /ɔ/, and a rating of the spectrographic quality of liquids and nasals. The seminal work of Monsen (1978) ultimately led us to consider seriously the development of an intelligibility test based on acoustic parameters that would have generalized explanatory power regarding the underlying dimensions of common surface level errors. A major problem confronting us, however, was that both Monsen (1978) and Stromberg and Levitt (1979) observed high intercorrelations among their predictor variables which render redundancy in the prediction process. The presence of high intercorrelations among predictor variables implies that many of the variables reflect the operation of a smaller set of dimensions that may relate to speech intelligibility at a more general and fundamental level. In order to circumvent the problem of high intercorrelations among acoustic predictor variables, we (Metz et al., 1985) employed a factor analytic procedure which would derive a smaller set of uncorrelated factors that could be used in lieu of the larger set of original acoustic variables in regression equations.

Metz et al. (1985) employed 10 original acoustic parameters that included segmental variables related to voicing contrasts and their production variability, vowel formant separations and their production variability, and second formant trajectory changes during productions of the diphthong /ai/ and its production variability. A simple suprasegmental variable, mean sentence duration and its production variability were the final two acoustic variables. Pure tone average was a non-speech independent predictor variable. When the original 10 acoustic variables were regressed on to estimates of transcription measures of speech intelligibility a mean VOT difference and the mean sentence duration variable accounted for 89% of the variance in the intelligibility estimates. The factor analytic procedure, however, derived simple structure from the original 10 predictor variables resulting in four factors. One of these factors (Factor 1) was composed almost exclusively of segmental variables and the second factor (Factor 2) was composed almost exclusively of the suprasegmental variables. When the factors were regressed on to estimates of transcription measures of speech intelligibility, Factors 1, 2, and 3 (Factor 3 comprised moderately low loadings of both segmental and suprasegmental variables) accounted for 92% of the variance in the intelligibility estimates. Importantly there was a clear separation of the segmental and suprasegmental components in the factor structure reflecting independent contributions of segmental and suprasegmental processes in the determination of speech intelligibility (i.e., Factors 1 and 2). The dominant factor in this prediction was Factor 1 thus confirming Levitt's (1971) observations on the clustering of speech production errors that influence intelligibility and quality.

The broad range of relatively pure segmental distinction variable loadings on Factor 1 and the predictive relationship of Factor 1 to our intelligibility measure provided empirical motivation for the theoretically reasonable assumption that control of spatial/temporal segmental events is a general and primary dimension affecting speech intelligibility. Importantly, inspection of the factor loading patterns (interpreted in much the same way one interprets bivariate correlation coefficients) of the original acoustic variables provides direct reference to the importance of that variable to the factor, and as such, to the deterioration of speech intelligibility. These findings were replicated and extended by Metz et al. (1990) thus forming the basis of the explanatory function of intelligibility testing that we seek.

Considered collectively, the findings of Monsen (1978), Metz et al. (1985), and Metz et al. (1990) strongly suggest the primacy of control over spatial/temporal segmental events in the determination of speech intelligibility among deaf persons. The strong predictive relationship between groupings of acoustic variables and speech intelligibility provided sufficient motivation for a preliminary attempt to develop an automated computer-based speech intelligibility estimation procedure that employs artificial neural networks (ANNs). Given the demonstrated power of computerized neural networks in speech recognition research, it is reasonable to suggest that accurate speech intelligibility estimates could be obtained by training an ANN to recognize patterns within groupings of selected acoustic variables that are associated with varying levels of speech intelligibility. Such predictive speech intelligibility estimates would not require active listener participation, and would possess the clinically useful attribute of having the intelligibility deficit described in terms of its underlying articulatory/acoustic dimensions.

Metz, Schiavetti, and Knight (1992) trained a back-propagation artificial neural network to recognize underlying patterns of acoustic speech parameters that are associated with varying levels of speech intelligibility. The parameters we employed to train the ANN were a subset of acoustic dimensions that had high factor loadings and a strong predictive association with measures of speech intelligibility. When the ANN was presented with sets of acoustic parameters derived from samples of speech produced by a different group of deaf persons, the derived speech intelligibility estimates exhibited good congruence with intelligibility estimates obtained from listeners. This retrospective pilot study has led to the development of a prospective experiment currently in progress to test the clinical tractibility of ANNs for the prediction of speech intelligibility from acoustic parameters.

CONCLUSION

Perhaps the most appropriate approach to suggesting measurement goals for evaluating remediation effectiveness is to specify two sets of criteria for clinicians/researchers to consider. First, the measures should be clear reflections of the physiological/acoustic/perceptual chain as it relates to speech intelligibility. That is to say, the measures should provide some link between what the speaker does with the articulators to the physical result and to the perceptual result. Second, the specific measurements targeted should have some or all of the following practical advantages. They should be objectively and reliably measured. Physiological and acoustic variables should be instrumentally measured and archived. Measures should be sensitive to the kind of small, but incremental changes often seen in remediation programs. Instrumental measures should be displayable for use in feedback to speakers during intervention, preferably online. Finally, measures should be obtained easily and economically so that they can be incorporated into intervention programs.

ACKNOWLEDGEMENT

This work was conducted under the course of an agreement between the United States Department of Education and the National Technical Institute for the Deaf, a college of Rochester Institute of Technology.

REFERENCES

Beiter, A., Staller, S., & Dowell, R. (1991). Evaluation and device programming in children. *Ear and Hearing, 12,* 25S-33S.

Boothroyd, A. (1985). Evaluation of speech production of the hearing impaired: Some benefits of forced-choice testing. *Journal of Speech and Hearing Research, 28,* 185-196.

Carney, A. (1986). Understanding speech intelligibility in the hearing-impaired. *Topics in Language Disorders, 6*(3), 47-59.

Davis, H. & Silverman, R. (1978). *Hearing and deafness.* New York: Holt, Rinehart, & Winston.

Hudgins, C., & Numbers, F. (1942). An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs, 25,* 289-392.

Kelly, C., Dancer, J., & Bradley, R. (1986). Correlation of the SPINE test scores to ratings of speech intelligibility in hearing-impaired children. *The Volta Review, 88,* 145-150.

Kent, R. (1992). *Intelligibility in speech disorders.* Philadelphia, PA: John Benjamins Publishing Co.

Kent, R., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Research, 54,* 482-499.

Levitt, H. (1971). Speech production for the deaf child. In L.E. Connor (Ed.), *Speech for the deaf child: Knowledge and use* (pp. 59-83). Washington, DC: A.G. Bell Association for the Deaf.

Ling, D. (1976). *Speech and the hearing-impaired: Theory and practice.* Washington, DC: A.G. Bell Association for the Deaf.

Metz, D.E., Samar, V.J., Schiavetti, N., Sitler, R., & Whitehead, R.L. (1985). Acoustic dimensions of hearing-impaired speakers' intelligibility. *Journal of Speech and Hearing Research, 28,* 345-355.

Metz, D.E., Schiavetti, N., & Knight, S. (1992). The use of artificial neural networks to estimate speech intelligibility from acoustic variables: A preliminary analysis. *Journal of Communication Disorders, 25,* 43-53.

Metz, D.E., Schiavetti, N., Samar, V.J., & Sitler, R. (1990). Acoustic dimensions of hearing-impaired speakers' intelligibility: Segmental and suprasegmental characteristics. *Journal of Speech and Hearing Research, 33,* 476-487.

Monsen, R. (1978). Toward measuring how well hearing impaired children speak. *Journal of Speech and Hearing Research, 21,* 197-219.

Monsen, R. (1981). A usable test of speech intelligibility of hearing-impaired talkers. *American Annals of the Deaf, 126,* 845-852.

Nickerson, R. (1975). Characteristics of the speech of deaf persons. *The Volta Review, 77,* 342-362.

Nickerson, R., & Stevens, K. (1980). Approaches to the study of the relationship between speech intelligibility assessments of hearing-impaired adults. In J. Subtelny (Ed.), *Speech assessment and speech improvement for the hearing-impaired* (pp. 338-364). Washington, DC: A.G. Bell Association for the Deaf.

Padden, C., & Humphries, T. (1988). *Deaf in America voices from a culture.* Cambridge: Harvard University Press.

Parkhurst, B., & Levitt, H. (1978). The effect of selected prosodic errors on the intelligibility of deaf speech. *Journal of Communication Disorders, 11,* 249-256.

Samar, V.J., & Metz, D.E. (1990). Criterion validity of intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research, 31,* 307-316.

Samar, V.J., & Metz, D.E. (1991). Scaling and transcription measures of intelligibility for populations with disordered speech: Where's the beef? *Journal of Speech and Hearing Research, 34,* 697-704.

Schiavetti, N., Metz, D.E., & Sitler, R. (1981). Construct validity of direct magnitude estimation and interval scaling of speech intelligibility. *Journal of Speech and Hearing Research, 24,* 441-445.

Shaw, S., & Coggins, T. (1991). Interobserver reliability using the Phonetic Level Evaluation with severely and profoundly hearing-impaired children. *Journal of Speech and Hearing Research, 34,* 989-999.

Smith, C. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research, 18,* 795-811.

Stevens, S.S. (1975). *Psychophysics.* New York: Wiley.

Stromberg, H., & Levitt, H. (1979, November). *Multiple linear regression analysis of errors in deaf speech.* Paper presented at the 97th meeting of the Acoustical Society of America, Cambridge, MA.

Subtelny, J. (1977). Assessment of speech with implications for training. In F. Bess (Ed.), *Childhood deafness* (pp. 183-194). New York: Grune and Stratton.

Tobey, E., & Hasenstab, S. (1991). Effects of nucleus multichannel cochlear implant upon speech production in children. *Ear and Hearing, 12,* 48S-54S.

Tye-Murray, N., & Kirk, K. (in press). Relationship between phonetic level evaluation and spontaneous speech: Vowel and diphthong production by young cochlear implant users. *Journal of Speech and Hearing Research.*

Weismer, G., Kent, R., Hodge, M., & Martin, R. (1988). The acoustic signature for intelligibility test words. *Journal of the Acoustical Society of America, 84,* 1281-1291.

Woodward, J. (1972). Implications for sociolinguistics research among the deaf. *Sign Language Studies, 1,* 1-7.

Woelfel, J., & Fink, E. (1980). *The measurement of communication processes.* New York: Academic Press.

Young, M. (1969). Observer agreement: Cumulative effects of rating many samples. *Journal of Speech and Hearing Research, 12,* 135-143.