1986

Validation of a speech intelligibility measure for profoundly deaf children

Cynthia J. Williams

Validation of a Speech Intelligibility Measure for Profoundly Deaf Children

Cynthia J. Williams

Sponsored by Ann Geers
Central Institute for the Deaf
Washington University
May 1986

I. INTRODUCTION

The practice of Deaf Education has undergone many changes in it's development since it's conception during the 18th century. Professionals of today are constantly evaluating and developing new techniques for teaching, helpful devices, and instruments of measurement for the various aspects of the hearing-impaired person. However, there is one specific area which has often been overlooked in development of quantitative tests. This area deals with the measurement of speech intelligibility of the hearing-impaired. The purpose of this study is to validate a recently developed measurement of speech intelligibility for profoundly deaf children.

The measurement of speech intelligibility has often been considered a strictly subjective value judgement on the part of teachers and clinicians. However, the usefulness of such a subjective measurement is somewhat limited. Meanwhile, new and more efficient tests were being developed to evaluate other areas, such as the aspects of hearing or language development. Many professionals felt a need for a time efficient, yet quality instrument for measuring speech intelligibility. This need was compounded by the underlying goal of all deaf education programs, be it an oral or total communication program, for good intelligible speech. This universal goal requires that speech be quantitatively monitored at appropriate intervals Afterall, one of the main reasons the educational needs are special for the hearing-impaired deals with the difficulty of teaching speech.

A system was developed to evaluate the intelligibility of speech. The subject was required to read a sentence or paragraph and the listener recorded what was understood. There are many drawbacks in utilizing this system. First of all, it is time-consuming and expensive. It requires a number of adult listeners in order to be objective. Furthermore, familar listeners may have an advantage, as well as pre-formed opinions of an individual's level of intelligibility. This system may also be a stressful situation for the subjects, and thereby influence their quality of speech. Additionally, it would be very difficult to be consistent. In order to utilize the scores for measurement of progress, the listeners must be equally capable in each measurement. Finally, the materials must always be carefully balanced. This pre-described method of evaluating speech intelligibility is rarely utilized because of it's difficulty in administrating and the time factor involved.

An alternate suggestion for measuring speech intelligibility has frequently been a phonetic analysis of a subject's speech. Monsen suggests that this is a difficult and impractical approach. One major problem arises in the training required by such an approach. Phonetics is an acquired skill. It takes years of training to achieve an expertise in the area. Furthermore, even after years of training, there are still variations from person to person. Also, phonetic analysis was originally designed to apply to "normal speech". The speech errors of the hearing-impaired are unlike those of the hearing. If the errors were of equal value, the most intelligible speaker

would be the one with the least amount of errors. This theory does not hold true for the hearing-impaired. They produce unique sounds and mistakes. The speech intelligibility is affected only if the sounds are not recognizable. Although Monsen does not recommend phonetic analysis for measurement of speech intelligibility, he states it can be valuable for speech remediation (Monsen, 1981).

Faced with the dilemma of the absence of a quantitative index for speech intelligibility, Monsen began to research some possibilities. The first step was to outline a definition of speech intelligibility. He defined speech intelligibility of a hearing-impaired subject as "a percent of simple sentences (both declarative and interrogative) understood by both experienced and inexperienced listeners in the absence of visual or verbal context". Monsen outlined four major quidelines in developing a quantitative index for speech intelligibility. First, it should be easy for teachers and clinicians to administer. Secondly, it should not require large amounts of time to administer. Thirdly, the index should be reliable from administrator to administrator and from test to test. Finally, the index should measure the individual's intelligibility of spoken language.

The resulting product of Monsen's research was the Speech Intelligibility Evaluation, or the SPINE. This test was developed at Central Institute for the Deaf in 1981. The SPINE met all of Monsen's prescribed criteria. The test is based upon the ability of any native speaker of English to make a phonemic classification of spoken sounds. The choice of words were based

on acoustic studies of the speech of the deaf (Monsen, 1978). The test consists of ten decks of cards. Each of these ten decks contain equal number of cards, each imprinted with one of four different words. The four words are different for each deck. The objective is to determine which of the four words was uttered, rather than how well it was produced. The administration of this test results in a speech intelligibility percentage. An interpretation of these scores will be included later in this paper. The results of the Word SPINE were validated with respect to the intelligibility of running speech of the hearing-impaired. Monsen recorded each subject's production of a set of sentences. These sentences were scrambled and played for a number of listeners. He found a high, positive correlation of .86 between the Word SPINE scores and the sentence intelligibility scores.

Upon completing the research for the Word SPINE, Monsen began developing a picture-only version for use with younger children. This test was under development in the fall of 1985 at Central Institute for the Deaf. The test is entitled. "The Photo-SPINE Test: A SPeech INtelligibility Evaluation for Hearing-Impaired Children". This study was undertaken to determine the validity of the Photo-SPINE test as a reliable predictor of the general intelligibility of connected speech.

II. METHOD

Subjects

The subjects aged 10-1 to 14-8, four boys and six girls,

were selected from among the school population at Central Institute for the Deaf. Each subject has been classified as having a profound hearing loss. In other words, their pure-tone audiometric average for .5, 1, and 2 kHz is greater that 95 dB in their better ear. The ten subjects were selected to represent, in the opinion of their teachers, a range of speech intelligibility from very good to very poor.

Materials

The Photo-SPINE test materials consists of a user manual, response forms, four charts illustrating all the pictures in each word set, and four "decks" of testing cards. The test pictures consists of four sets of 25 minimally different words. Each set contrasts a specific group of sounds. The words and features measured are listed in Table 1. The test words had to be nouns that were easily represented in picture form. This test is based strictly on spoken responses to pictures rather than to written words, as in the Word-SPINE. The reading task was completely eliminated from this test.

Procedure

The Photo-SPINE was administered according to the prescribed procedure in the user's manual (Monsen, 1985). This procedure is as follows:

1) The subject to be evaluated and the examiner are seated across from each other at a table in a quiet room. The examiner is allowed to utilize visual and auditory skills to score the subject's response.

2) The examiner presents the practice cards for Set 1. This

Table 1: The Four Sets of Words Used in the Photo-SPINE and the Features of Speech which they Represent.

Test Words						Features Measured
Set 1:	pea	bell	bed	pen	book	labial initial
	bee	ball	bird	pin	back	consonants v.s.
	bear	bowl	board	pan	bag	non-labial final
	pear	belt	bat	bean	bug	consonants
	pie	pail	boy	barn	pig	
Set 2:	doll	lake	tee	net	tail	alveolar initial
	door	leg	dee	nut	nail	consonants v.s.
	dog	lock	knee	nest	tongue	non-labial final
	duck	neck	tree	knot	toe	consonants
	deer	log	ten	nose	tow	
Set 3:	hat	hair	car	cap	goat	velar initial
	heart	horse	card	cup	coat	consonants v.s.
	hand	house	cow	cake	kite	alveolar final
	head	hose	cat	eight	gate	consonants
	eye	ear	key	chair	girl	
Set 4:	sheep	seal	rug	gun	rain	mixed initial
	ship	shell	rock	gum	ring	consonants v.s.
	soap	saw	rake	comb	can	mixed final
	soup	wing	robe	corn	game	consonants
	sun	one	rope	horn	clown	

set has one picture each of the twenty-five words. The subject rehearses the correct name for each picture. The examiner determines whether the words are familiar to the subject. The test can not be administered if the words are unfamiliar.

3) The examiner explains the nature of the test to the subject. That is, the subject is expected to pronounce the word pictured on each card shown to him. There are 50 cards in each test set, two of each picture. The examiner then tries to determine which word, from among the twenty-five possible choices, was said.

4) After familarizing the subject with the procedure, the examiner removes the test deck for set 1, shuffles them to avoid any chance of pattern, and may begin the testing. The examiner continues in like manner until all four sets have been completed. It is important to return the cards in the exact order for scoring later. A total of 200 words will be scored, resulting in a .5% value for each word. The resulting scores are listed in Table 2.

Monsen outlined a rating scale to help interpret the percentage scores in terms of intelligibility (Monsen, 1981). The scores can be interpreted in the following terms:

100-90% correct: excellent achievement in speech intelligibility. Naive listeners can understand most of the child's speech at first introduction.

-80% correct: good achievement in speech intelligibility. Naive listenerss will "miss" occasional words in sentences,

but most will be understood, and the communication process on the whole is smooth.

79-70% correct: Listeners experience difficulty in understanding the intended message. But communication can nevertheless take place, though with noticeable difficulty.

69-60% correct: Listeners experience great difficulty in understanding simple material. The communication process is labored and difficult.

59% and below: Listeners are confronted with overwhelming difficulty in understanding what was said. Only occasional words can be picked out of the flow of speech. Even experienced listeners will experience great difficulty understanding what was said. Communication begins to center upon the conveying of nouns, often by means of gestures.

Table 2: Subjects' Speech Intelligibility Scores from the Photo-SPINE and the Sentence Imitation Task.

Subject Number	Photo-SPINE Score (in percent)	Mean Intelligibility Score (in percent)
1	99	99.5
2	97	95.95
3	96	97.55
4	95	94.8
5	79	51.3
6	68	47.25
7	66	63.5
8	65	51.75
9	64	42.4
10	54	53.65

In addition to the Photo-SPINE scores, the speech intelligibility of the same hearing-impaired subjects was subsequently measured in a different manner. Sentences said by the subjects were placed on listening tapes and presented to adult listeners. The sentences are listed in Appendix A.

Each subject was asked to imitate ten sentences each. The examiner prepared two listening tapes, 50 sentences each. The tapes contained two identical, immediately successive repetitions of the sentences. The order of subjects and sentences was randomized for each tape. These tapes were

presented to ten experienced listeners. That is, they all work with the hearing-impaired in some capacity. Each listener was instructed to write down in normal English orthography what he thought each subject said. They were allowed to guess if necessary, but asked not to write down phonetically what the subject's utterance "sounded like". The resulting intelligibility scores are based on the listeners' judgements of 1000 sentences (10 listeners X 10 subjects X 10 sentences).

In scoring the responses, each sentence was accorded a maximum value of 10, and words within the sentence were assigned individual point values depending on their frequency in the language. The higher the frequency and predictability of a word w[illegible], the lower the assigned value. For example, the words in the sentence, "I like ice cream" were scored 1, 2, and 7 points respectively. No points were awarded for partially correct words. The listeners' responses for each subject resulted in a percentage intelligibility score. These scores ranged from 100% correct to 11.5% correct. These scores are listed in Table 2. The mean intelligibility score was 69.8%

III. RESULTS AND DISCUSSION

The two sets of intelligibility scores were compared by the Pearson's r coefficient of correlation. A very high positive correlation was obtained: +.82. This high degree of relatedness between the two sets of scores indicates that the Photo-SPINE Test is, in fact, highly predictive of speech intelligibility for profoundly deaf children.

These results provide evidence for the validity and usefulness of the Photo-SPINE Test in measuring the speech intelligibility of hearing-impaired subjects. This test has a variety of practical uses to the clinician or classroom teacher. First, an individual's speech intelligibility progress may be monitored by a comparision of scores over a period of time. A student's progress or lack of progress can addressed and dealt with appropriately. Parents are also likely to appreciate some objective feedback concerning the intelligibility of their child. Secondly, different methods of teaching speech can be evaluated and compared utilizing the Photo-SPINE scores. Thirdly, a child's intelligibility score may aid administrators' decisions concerning appropriate educational placement for that child. For example, a child with a score of 65% may not be considered a candidate for a mainstream situation. Finally, specific techniques for teaching speech may be evaluated with a valid tool to measure their successfulness.

In conclusion, this study has shown that the Photo-SPINE Test is a valid instrument with which to measure the speech intelligibility of the profoundly deaf child. This time-conscious test is easy to score. Furthermore, this test is more appropriate than the Word-SPINE for younger children because the reading task has been eliminated. This factor may also be important when dealing with children who may have visual and/or learning problems. The measurement of speech intelligibility has finally reached the level of sophistication present in other areas of evaluation of the hearing-impaired.

References

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

Monsen, R.B. A usable test for the speech intelligibility of deaf talkers. American Annals of the Deaf, 1981, 126, 845-852.

Monsen, R.B. The oral speech intelligibility of hearing-impaired talkers. Journal of Speech and Hearing Disorders, 1983, 48, 286-296.

Appendix

Appendix A

Set A

1. The ice cream was good.
2. He plays baseball.
3. He has a lunchbox.
4. The toothbrush is red.
5. We eat lunch at noon.
6. We all like sports.
7. I don't think so.
8. Did you find it?
9. Can they hear it?
10. Can you tell us?

Set B

1. The football is brown.
2. I want a hot dog.
3. They have an airplane.
4. The ice box is full.
5. I know how to dance.
6. We bought a new car.
7. Let me read this.
8. Will he like them?
9. Do you know how?
10. Can I come in?

Set C

1. The playground is clean.
2. I like ice cream.
3. I need a toothbrush.
4. The lunchbox is small.
5. Please help me do it.
6. Our baseball is lost.
7. I don't want any.
8. Will you go there?
9. Will he see it?
10. Can you show me?

Scores and Averages from Sentence Imitation Task

(Listeners)

Subjects

	AI	MF	AK	MF	JE	PD	KK	LK	LM	LL	Average	Spine	
JK	[illegible]	44	42.5	57	24	45.5	57	47	61	[illegible]	47.25	65	
GB	[illegible]	42	61.5	30	34	54	51.5	95.5	[illegible]	32.5	53.65	54	
CC	100	91.5	100	85	71	100	[illegible]	96.5	[illegible]	90	95.95	91	
SW	42.5	42.5	41.5	24	11.5	36.5	54.5	50	73	37	42.4	64	
DM	100	91.5	100	[illegible]	100	100	100	100	100	100	99.95	99	
SR	50	40.5	45.5	42.5	25.5	[illegible]	56	69	[illegible]	44	51.3	79	
PT	100	90	100	100	71	100	91	99	100	85.5	97.55	96	
DL	69	56	45	60.5	49	61.5	50	90.5	87.5	53.5	63.5	66	
LS	100	80.5	95.5	75	93	100	100	[illegible]	96	100	94.8	95	
KK	69	49	32.5	67.5	[illegible]	41.5	57	57	64	49	51.75	65	

$\bar{x} = 69.8$ $\bar{y} = 78.3$

S.D. = 24.1 S.D = 16.79

Pearson R = .82