1968

An Investigation of Scaled Severity of Whistled [s]

Mary Fortner Altman
Eastern Illinois University

PAPER CERTIFICATE #3

To: Graduate Degree Candidates who have written formal theses.

Subject: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements.

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

Aug 1, 1968
Date Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because _____

_____

_____

_____ _____
Date Author

AN INVESTIGATION OF SCALED SEVERITY OF WHISTLED [s]

(TITLE)

BY

MARY FORTNER ALTMAN

B.S. in Ed., Eastern Illinois University, 1966

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1968
YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

August 1, 1968
DATE

Aug 1, 1968
DATE

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The author expresses her thanks to Dr. Jerry Griffith for serving as an adviser for this thesis and also for serving as one of the judges, and to Dr. Wayne L. Thurman for his help as an academic adviser and a judge for this thesis. Other judges to whom gratitude is extended are Dr. James M. Plugrath, Mr. Lynn Miner, Mrs. Viola Hallock, Miss Bonnie Barnett, Mrs. Margaret Barlow, Mr. Blaise DeMuth, Mr. George Dudley, and Miss Phyllis Kinkade.

A special thank you goes to Mr. Lynn Miner and Mr. Bill Elrick for their assistance at Data Processing with the computer.

To many other friends, fellow students and family who have provided encouragement and assistance in numerous ways the author says thanks.

CHAPTER I

INTRODUCTION

The whistled [s] has been debated and discussed by speech pathologists for years. Speech clinics and public school therapists often have caseloads partially comprised of persons who misarticulate the /s/ phoneme in such a manner that we label it whistled. Even so, it seems that a whistled [s] means different things to different speech pathologists.

Therefore, it appeared to be worthwhile to isolate the whistled [s] and rate it on a severity scale. If a recorded severity scale of whistled [s] were available with examples of the severity of distortion for each interval on the scale, potential judges could listen to a randomized tape of these responses and rate the distortions of [s] on a nine-point equal-appearing interval scale. If their judgements agreed significantly with the previously established severities of the samples, then they might be considered reliable judges on similar tasks.

If we are going to include whistled [s] distortions in our case loads, then it is important to know with what we are dealing. How far from a normal [s] is this whistled [s]? In what phonetic context is this person's production of [s] less severe or more severe?

A recorded severity scale could be used in training therapists, providing examples of degrees of severity of the whistled [s].

This study was designed in expectation of providing such a scale and to investigate the effects of selected phonetic contexts on the rated

severity of [s] distortions commonly described as whistled. For the purpose of this study a whistled [s] is defined as an /s/ production which is accompanied by flow of the constricted breath stream across the edges of the central incisors in such a fashion as to impart a near-tonal quality to the resultant sound. The method of this investigation was to obtain a sample of whistled [s] in each of as many specified phonetic environments in the English language as possible and to present these samples of [s] to a panel of judges who were asked to rate each of the samples on a nine-point equal-appearing interval scale. From these data an attempt was made to compose a tape with representative utterances of whistled [s] at each of the nine equal-appearing intervals. Specifically, the study was designed to provide answers to the following questions:

1. Can experienced speech pathologists agree on the severity of utterances of whistled [s] using a nine-point equal-appearing interval scale?

2. Is it possible to compile a tape recorded scale of speech responses for the purpose of determining the reliability of judgements for any given listener from the original utterances of the whistled [s] allophone judged as belonging to one of nine points on a nine-point equal-appearing interval scale?

3. Do whistled [s] distortions occur more frequently in certain phonetic contexts than in others?

CHAPTER II

REVIEW OF THE LITERATURE

There have been many tests and scales devised to measure articulation severity. These have been efforts to quantify or give a value to articulation severity. Wood (1949) weighted all sounds as portions of 100% depending on Travis' (1931) count of frequency of sounds in the English language. He then judged articulation severity according to the score obtained by subtracting total weighted scores for all errors from 100.

Templin (1953) tested the reliability of a 50-item screening test against a 176-item test and found the screening test reliable above .93.

Reid (1947) based an articulation defective scale on a reverse developmental order giving the earliest learned sounds the highest numbers and subtracting the error total from the total possible score. Curry, Kennedy, Wagner and Wilke (1943) constructed a phonographic rating scale for measuring defective articulation by having observers judge recorded samples of speech ranging from nearly inarticulate speech to normal speech using a paired-comparison technique. The use of their scale requires that the experimenter must first familiarize himself thoroughly with the recorded samples of defective speech and it is recommended that he systematically refer to the scale steps as a check on accuracy.

Wright (1954) saw the need for ". . . a finer measurement tool for defective articulation . . . than the customary recording of sounds as correct, distorted, substituted, or omitted (p. 21)." He used a seven point recording scale to determine the reliability of a judge's evaluations

made during basic articulation and stimulation testing. This scale began with a correctly articulated sound, followed by four levels of distortion, then substitution and omission, recognizing ". . . the previously neglected fact that a defective sound may vary in degree of distortion, as well as being substituted or omitted (p. 21)." Wright concluded that reliable evaluations can be made during articulation and stimulation testing according to this procedure.

Other studies, such as those of Morrison (1955), Jordan (1960), and Lewis and Sherman (1951), reported the use of a nine-point severity scale. Cullinan, Prather and Williams (1963) found that five, seven and nine-point scales seem to be equally good while the interjudge reliability on the direct magnitude-estimation was considerably lower when less than ten judges were used. The direct magnitude-estimation yields a ratio scale.

> In this procedure a sample of approximately average severity, referred to as the standard sample, is presented to the judges. Each judge is free to assign whatever number he wishes to represent the severity of this sample. The experimental samples are then presented to the judges who assign numbers which represent the severity of each sample relative to the standard sample (pp. 188-189).

Sherman and Moodie (1957) found that of equal-appearing intervals, successive intervals, paired comparisons and constant sums methods of scaling articulation defectiveness from short segments of speech, the equal-appearing intervals technique seemed to be the most useful:

> In view of the demonstrated reliability of scale values obtained by the method of equal-appearing intervals and their close agreement with the internally consistent scale values obtained by the method of successive intervals, the method of equal-appearing intervals would, in general, be the preferred choice for the task of scaling short segments of speech with respect to articulation defectiveness. The simplicity of the method, including computational procedures, makes possible the scaling of many more speech segments with a reasonable amount of time and labor than either of the other two methods, particularly the method of pair comparisons (p. 704).

In light of the findings of the Sherman and Moodie (1957) study it

was decided that the equal-appearing intervals scale would be best suited for the judging tasks of the present study; and the nine-point severity scale was chosen on the basis of the studies by Morrison (1955), Jordan (1960), and Lewis and Sherman (1951).

CHAPTER III

PROCEDURE

Subjects: Twenty university students (fifteen females, five males) served as subjects. Each subject had been diagnosed by a member of the staff at the Speech and Hearing Clinic, Eastern Illinois University, Charleston, Illinois as having a whistled [s] distortion and was currently receiving speech therapy for correction of this distortion. These subjects provided the responses used in the judgement procedures.

Selection of the Rating Scale: A nine-point equal-appearing intervals scale of severity with 1 representing a "normal" [s] and 9 representing "the most severe whistled [s] ever heard" was chosen on the basis of the Sherman and Moodie (1957) study cited earlier. The nine points were chosen because smaller interval scales did not seem to provide as reliable results, and the longer interval scales did not seem to increase reliability. It was also thought that the smaller the scale, the higher the percentage of agreement that would be required in order to be statistically significant, and the smaller the scale, the greater the risk of a mathematical artifact causing the agreement rather than actual judge agreement.

Selection of the Word Lists: A list of 154 words representing /s/ in the releasing position in both stressed and unstressed syllables and in all possible phonetic contexts found in the English language were chosen from the Speech Clinician's Handbook (Miner, 1968). It was also required that the chosen words appear in The Teacher's Word Book (Thorndike, 1921). The tested sound, /s/, occurred only once in any test word. This reduced

confusions as to which sound was being evaluated. Syllabification was checked in either Webster (1956) or Kenyon and Knott (1953). The word lists are found in Appendix I.

Selection of Judges: Ten judges were selected according to the following criteria: 1) at least one year of experience as a speech clinician; 2) at least beginning work toward a Master's degree in speech pathology or with emphasis in speech pathology; 3) normal hearing as determined by a screening audiometric evaluation at 25 dB (I.S.O. 1961) from 250 to 8,000 Hz. One judge failed to meet criteria in that she could not hear 8,000 Hz. at a 25 dB level. She was included in the study because it was felt that this did not constitute a significant hearing loss. These evaluations were made on a Beltone 15C audiometer in a sound-proof room (Industrial Acoustics Company, Incorporated).

Preparation of tapes: Each subject spoke each of the 154 words while seated in a sound proof room. A distance of eight to ten inches between the speaker's lips and the microphone was maintained. The responses were monitored on a VU meter and the recorder level adjusted so that the responses peaked at an average of 0. Responses were recorded on an Ampex Recorder/Reproducer, Model 602, with an Astatic Microphone, Model 77A, on Mylar Scotch 190 recording tape. Stimulus cards had been prepared by using a primary typewriter centering the words on white five by seven inch cards. Each card was presented to the subjects at timed intervals so that the words were spoken at the rate of one word every three seconds to allow for splicing.

The foregoing procedure resulted in a sample of 3,080 responses to be used for judging. The tape recorded responses were then numbered with a felt tip marker for the purpose of identification of speakers and words.

The tapes were then cut between words and spliced together in random order at two second intervals for playback. The two second interval was chosen on the basis of Morrison's (1955) study in which she successfully used two second latency periods for judging. A 155 page answer sheet was then prepared to correspond to the randomized order of the responses so that the data could be transferred to IBM cards and processed by the computer. The scale numbers alternately ran from one to nine, then nine to one so as to avoid a constant motor tendency to check at one side of the page (Guilford, 1936). A sample answer sheet is included as Appendix II.

Judgement Procedure: The tape recorded responses were presented to the panel of ten judges in a sound-proof suite. From one to five judges listened and judged at the same time by using a jack hook-up between two sound-proof rooms. Each judge listened through Telex MR-6 earphones. Each judging session lasted approximately two hours on each of two different days for a total of four hours per judge. One person judged on three different days because of circumstances beyond our control. This was not considered a serious variation. Two or three minute breaks were taken every fifteen minutes during the judging sessions.

Before each session the judges listened to and read taped instructions and judged a practice tape of ten words. These ten worda were not included in the analyses nor did they appear in the test tapes. The instructions are shown in Appendix III.

The judges were asked to rate each response on a nine-point equal-appearing interval scale of severity with 1 representing a "normal" [s] and 9 "the most severe <u>whistled</u> [s] ever heard."

Statistical Analysis: A total of 30,800 judgements were then

available for analyses. The judges' answers were transferred from the answer sheets to IBM cards so statistical analyses could be performed by the computer. All statistical analyses were done on an IBM 1620 computer.

The following statistical analyses and comparisons were made:

1. A mean scale judgement rating for each word for each speaker.
2. A measure of variability of scale judgements in terms of a standard deviation for each word for each speaker for each judge.
3. The percentage agreement among judges for each word for each speaker.
4. A comparison of mean ratings for each speaker for each word.
5. A measure of the effect of phonetic context on the degree of distortion of /s/ for each speaker by determining the average scale values for each word for all speakers.
6. A mean rating and standard deviation for each word for each speaker for each judge.
7. Percentage agreement among judges for each word for each speaker.

CHAPTER IV

RESULTS

Table 1 shows the 19 out of 3,080 utterances on which at least seven of the ten judges agreed in their ratings on the scale. The speakers are identified by number. Speakers 19, 7 and 11 each uttered one word on which there was 70% or greater agreement by the judges on scale value. After a review of the literature and discussion it was arbitrarily decided that 70% agreement among the judges would be appropriate. Speakers 6 and 5 uttered two significant words, speaker 9 uttered three and speaker 4 uttered nine significant words, if we use 70% agreement to define significant. One word "extract," appears twice in the nineteen significant utterances.

The judges' use of the scale intervals is indicated as a total value under each interval of the nine-point equal-appearing intervals scale. The total frequency of use of each interval was determined by summing the values under each interval and is indicated at the bottom of the table as "Total frequency of use." The total frequency of use of each interval for all 154 words was determined by summing the judgements in each interval for all 154 words and is shown as "Total frequency of use for 154 words."

The mean and standard deviation of judgements of the scale values for each word is indicated on the same line as that word. Words 1, 5, 8, 9, 10, 11, and 14 have standard deviations of 1.0 or less.

Table 2 lists the same nineteen words as in Table 1 and shows each

TABLE 1.--Summary of words reaching 70% agreement criterion showing speaker number, word, distribution of judgements, and means and standard deviations of judgements. Frequency of use of each scale interval for these words and all 154 words is shown for purposes of comparison.

	Speaker # Word #	WORD	1	2	3	4	5	6	7	8	9	$\bar{M}$	S.D.
1.	9-80	extract	7	2	0	1	0	0	0	0	0	1.5	.92
2.	4-90	astronomer	8	1	0	0	1	0	0	0	0	1.5	1.2
3.	9-46	wholesale	7	1	1	0	1	0	0	0	0	1.7	1.3
4.	4-47	landscape	7	1	0	0	1	1	0	0	0	2.0	1.8
5.	4-33	manuscript	7	3	0	0	0	0	0	0	0	1.3	.46
6.	4-101	swore	7	2	0	0	1	0	0	0	0	1.6	1.2
7.	4-95	spoke	8	0	1	0	0	1	0	0	0	1.7	1.5
8.	19-26	scripture	0	7	1	1	1	0	0	0	0	2.6	1.0
9.	6-51	schedule	7	1	1	1	0	0	0	0	0	1.6	1.0
10.	7-116	swarm	2	7	1	0	0	0	0	0	0	1.9	.54
11.	4-150	inspiration	7	2	0	1	0	0	0	0	0	1.5	.92
12.	9-122	manslaughter	7	0	1	2	0	0	0	0	0	1.8	1.2
13.	4-78	Scandinavian	7	0	1	1	1	0	0	0	0	1.9	1.4
14.	4-154	catastrophe	7	1	1	1	0	0	0	0	0	1.6	1.0
15.	15-40	snail	7	0	2	0	1	0	0	0	0	1.8	1.3
16.	11-81	sergeant	7	0	1	1	0	1	0	0	0	2.0	1.7
17.	6-80	extract	7	1	1	0	0	1	0	0	0	1.8	1.5
18.	15-37	escape	0	1	7	1	0	0	1	0	0	3.4	1.3
19.	4-76	strap	7	0	1	2	0	0	0	0	0	1.8	1.2
Total frequency of use			116	30	20	12	7	4	1	0	0		
Total frequency of use for 154 words			5839	5509	5413	3479	2602	2980	2566	1471	928		

TABLE 2.--Summary of scale values assigned by each judge for each word reaching the 70% agreement criterion. The sum of the scale values assigned by each judge for these words and all 154 words is shown for purposes of comparison.

	WORD	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10
1.	extract	1	1	2	1	1	1	2	1	1	4
2.	astronomer	5	1	2	1	1	1	1	1	1	1
3.	wholesale	1	5	1	2	1	1	1	1	3	1
4.	landscape	1	1	6	2	1	1	1	1	1	5
5.	manuscript	1	1	2	2	1	1	1	1	1	2
6.	swore	1	1	5	2	1	1	2	1	1	1
7.	spoke	1	1	6	3	1	1	1	1	1	1
8.	scripture	2	3	2	2	5	2	4	2	2	2
9.	schedule	1	1	4	3	1	2	1	1	1	1
10.	swarm	1	2	3	2	2	2	2	2	2	1
11.	inspiration	1	2	4	1	1	1	1	1	1	2
12.	manslaughter	1	4	3	1	1	1	1	1	4	1
13.	Scandinavian	1	3	5	1	1	1	4	1	1	1
14.	catastrophe	1	1	3	1	1	1	4	1	2	1
15.	snail	1	1	3	1	1	3	5	1	1	1
16.	sergeant	1	1	4	1	1	3	1	1	6	1
17.	extract	1	1	1	1	1	2	3	1	6	1
18.	escape	2	3	2	2	2	2	2	1	7	3
19.	strap	1	1	3	1	1	4	1	1	1	4
Sum of significant words		26	34	62	31	26	32	39	22	43	34
Sum of 154 words		10818	13562	12439	8773	14650	10119	13007	5923	15952	10735

judge's rating of each word. The sum of the ratings for all nineteen words for each judge is indicated at the bottom of each column. In addition the sum of ratings for each judge for all 154 words used in the study is shown. The sums for the significant words and for all 154 words indicate whether a judge tended to rate low or high overall. Placing the judges in rank order using only significant words we find Judge 8 rated the words least severe, followed by Judges 1 and 5, then Judge 4, Judge 6, Judges 2 and 10, Judge 7, Judge 9, and most severe Judge 3. By using the sum of 154 words and placing the judges in rank order from least severe to most severe we find: Judge 8, Judge 4, Judge 6, Judge 10, Judge 1, Judge 3, Judge 7, Judge 2, Judge 5 and Judge 9.

In Table 3 all 154 words are listed by number and are followed by the number of times each word received a rating with a variability of 1.0 standard deviation or less when uttered by each of the twenty speakers. It was arbitrarily decided that 1.0 standard deviation or less was appropriate criterion. An example of this criterion is shown where only four out of the twenty speakers uttered word number 1 and received a rating with a variability of 1.0 standard deviation or less. All remaining words may be examined in the same way.

Ninety-eight different words met this criterion a total of 148 times. This left fifty-six words which did not meet the criterion.

The eighteen words on which 70% of the judges made the same rating are noted with an asterisk. Two of these eighteen words had ratings that varied greater than 1.0 standard deviation.

TABLE 3.--Frequency with which each of 154 words received scale ratings with variability of 1.0 standard deviation or less. The words reaching the 70% agreement and shown in Tables 1 and 2 are followed by an asterisk.

WORD NO.	F	WORD NO.	F	WORD NO.	F	WORD NO.	F	WORD NO.	F	WORD NO.	F
1	4	27	0	53	0	79	2	105	0	131	1
2	0	28	1	54	0	80*	2	106	1	132	3
3	1	29	0	55	0	81*	2	107	1	133	0
4	0	30	0	56	1	82	1	108	0	134	0
5	0	31	2	57	0	83	0	109	1	135	3
6	1	32	0	58	3	84	0	110	1	136	0
7	0	33*	3	59	0	85	1	111	1	137	0
8	1	34	0	60	0	86	3	112	0	138	0
9	0	35	1	61	0	87	0	113	1	139	1
10	0	36	1	62	3	88	1	114	2	140	3
11	1	37*	2	63	0	89	1	115	0	141	0
12	0	38	2	64	2	90*	1	116*	2	142	1
13	2	39	0	65	3	91	1	117	1	143	1
14	1	40*	2	66	1	92	2	118	2	144	0
15	2	41	0	67	0	93	1	119	1	145	0
16	1	42	1	68	1	94	1	120	0	146	1
17	0	43	1	69	1	95*	1	121	0	147	1
18	4	44	1	70	0	96	3	122*	1	148	1
19	1	45	2	71	1	97	1	123	1	149	0
20	0	46*	1	72	2	98	0	124	1	150*	1
21	0	47*	0	73	2	99	1	125	2	151	1
22	0	48	1	74	1	100	1	126	2	152	1
23	1	49	2	75	1	101*	0	127	0	153	0
24	0	50	1	76*	3	102	1	128	1	154*	0
25	2	51*	1	77	0	103	3	129	0		
26*	1	52	2	78*	1	104	0	130	1		

* Words on which 70% of the judges made the same rating

148 utterances of 98 different words
56 words are not accounted for

CHAPTER V

DISCUSSION

The data from this investigation demonstrated that not 70% of a group of experienced speech pathologists could agree on the severity of recorded samples of whistled [s] using a nine-point equal-appearing interval scale. The following interpretations and generalizations might be made.

1) Only 19 utterances out of 3,080 reached the 70% agreement criterion (cf. Table 1). However, the probability that seven or more of ten judges would pick the same number on the scale by chance is approximately twice in 10,000 times ($p = .00002$) (Ferguson, 1959). The analysis shows that the judges exceeded this level significantly by agreeing six times in 1,000 ($p = .006$). In other words, the odds of getting the level of agreement achieved by the ten judges alone are 367:1.

2) Not all nine points on the scale were used when the 70% agreement criterion was met (cf. Table 1).

3) Table 2 shows that the judges did not retain their rank order of judged severity between the nineteen utterances and all judgements. For example, in comparing the sum of judgements for the nineteen utterances with the sum of all judgements we find that Judge 3 judged most severely for the nineteen utterances, but Judge 9 judged most severely for all judgements.

4) Table 3 shows the number of times utterances of each of the 154

words received judgements with variability of 1.0 standard deviation or less. There were 148 utterances of 98 words meeting this criterion. This accounts for only 64% of the original list. In addition, three of the words reaching the 70% criterion had standard deviations greater than 1.0. The remaining sixteen words were spoken only twenty-four times in such a way as to meet this criterion. The possible number of times they could have met criterion was 380. Therefore, this data does not show any consistent trends in favor of using a nine-point equal-appearing interval scale for this task. None of the speakers and none of the words seemed to contribute anything to the judgements beyond a chance level.

5) The order of frequency of the scale intervals was as follows:

Interval one was used 5,839 times or 19% of the time.

Interval two was used 5,509 times or 17.9% of the time.

Interval three was used 5,413 times or 17.6% of the time.

Interval four was used 3,479 times or 11.3% of the time.

Interval five was used 2,602 times or 8.5% of the time.

Interval six was used 2,980 times or 9.7% of the time.

Interval seven was used 2,566 times or 8.3% of the time.

Interval eight was used 1,471 times or 4.8% of the time.

Interval nine was used 928 times or 3% of the time.

Why were the results of this study not significant? Several hypotheses appear tenable.

At the onset of this study certain assumptions were made. One of these was that experienced speech pathologists could rate utterances of a <u>whistled</u> [s] on a nine-point equal-appearing interval scale. From the statistical evidence of this study one can see that they did not success-

fully agree on the severity of recorded whistled [s]'s. Why didn't they?

Perhaps the judges did not understand the task, even though recorded and printed instructions were given prior to each judging session. These instructions may not have been clear.

Possibly the judges were not able to discriminate between [s]'s. The whistled [s] was defined, but each judge may have had his own internal definition and used it unconsciously, or he may not have been consistent in his evaluation of what constitutes a whistled [s]. For instance, some authorities refer to certain /s/ distortions as "excessively sharp" or "excessively sibilant." There may possibly be a difference between the whistled [s] and an "excessively sharp" or "excessively sibilant" [s]. All whistled [s]'s are "excessively sharp" or "excessively sibilant," but all "excessively sharp" or "excessively sibilant" [s]'s are not necessarily whistled.

The judges may not have been consistent either within, or among themselves, in their use of the scale.

Fatigue of the judges could be a factor. Two hour judging sessions may have been too long. However, if this is true, it seems the results would have revealed a difference in judgements between the beginning and end of each judging session. No such difference was detected.

Perhaps the constraint of human memory is a factor. According to Miller (1965) ". . . the span of immediate memory impose(s) severe limitations on the amount of information that we are able to receive, process, and remember (p. 265)." He also says ". . . there is a finite span of immediate memory and that for a lot of different kinds of test materials this span is about seven items in length (p. 257)." In the process of

judging 3,080 utterances is it possible to remember the proper position that each particular utterance should be placed on a nine-point equal-appearing interval scale? Is it possible to remember the relative value assigned to each interval on the scale and can one store enough examples for each interval to be able to compare a given sample to previously judged samples?

Next, one might examine the scale. A nine-point scale has been reported as valid elsewhere (Jordan, 1960; Morrison, 1955; Lewis and Sherman, 1951; Cullinan, Prather, Williams, 1963), but did not work in this study. On each utterance that met the criterion of 70% or more agreement never more than seven points on the scale were used. Perhaps nine points were too many, or too refined for rating samples of whistled [s].

Lack of significant results obtained in this study might be caused by the speakers. Perhaps they did not represent a homogeneous clinical population in spite of the fact that they had been so labeled. It is conceivable that the inconsistency of their production of /s/ ranged from a broad or lateral [s] which might rightly have been judged on a nineteen point scale running from -9 to +9 with 0 being a "normal" production of /s/, -9 being a "most severe lateral distortion" and +9 being a "most severe whistled [s]." Maybe some of the utterances did not fit the description of a whistled [s] as defined in this study.

Another explanation for lack of significant results might include the procedure used in this study. The utterances were presented approximately one every two seconds. This may have been too fast for objective judgement in spite of the fact that this rate was chosen on the basis of Morrison's (1955) study previously cited. Two second latency periods may

have been sufficient for judging in the Morrison study, but may not have been sufficient for this task.

Both Morrison (1955) and Sherman and Moodie (1957) used tape recorded samples played at fifteen inches per second. This study used tape recorded samples played at seven-and-one-half inches per second. Perhaps the fidelity of this recording was not true enough to allow for the fine distinctions necessary on a nine-point scale.

According to Irwin (1965) the accurate identification of misarticulations was significantly better for audio-visual representations of sounds than for audio alone. If visual clues are important in diagnoais of general articulation errors, perhaps they are also important in judging severity ratings of the whistled [s].

CHAPTER VI

SUMMARY AND CONCLUSIONS

In this study three pertinent questions were asked at the outset: (1) Can experienced speech pathologists agree on the severity of whistled [s] utterances using a nine-point equal-appearing interval scale? The whistled [s] was defined as an /s/ production which is accompanied by flow of the constricted breath stream across the edges of the central incisors in such a fashion as to impart a near-tonal quality to the resultant sound. (2) Is it possible to compile a tape recorded severity scale of the whistled [s] with examples of distortions for each point on the scale? (3) Do whistled [s] distortions occur more frequently in certain phonetic contexts than in others?

The method of this investigation was to obtain a sample of misarticulated /s/ in specified phonetic environments and to present this sample of [s] to a panel of judges who were asked to rate each of the samples.

Twenty subjects, previously diagnosed as having a whistled [s] distortion each uttered 154 words representing /s/ in the releasing position in both stressed and unstressed syllables and in each of as many specified phonetic environments in the English language as possible. The resulting 3,080 responses were numbered, cut between words, and spliced together in random order at two second intervals.

Ten experienced speech pathologists judged the utterances on a nine-point equal-appearing interval scale. Statistical analyses of the 30,800 judgements were performed by computer.

The results showed that the judges agreed significantly (.70) on only 19 out of the 3,080 utterances. This means that, in answer to the first question, the judges could not agree on the severity of recorded whistled [s]'s using a nine-point equal-appearing interval scale. This data does not show any consistent trends in favor of using a nine-point equal-appearing interval scale for this task.

The answer to question two is also negative within the limits of this study. Because of the lack of significant agreement among judges for any one interval, it was not possible to compile a tape recorded severity scale of the whistled [s] with examples of distortions for any interval on the scale.

In answer to the third question, there were insufficient significant results relating to phonetic context to determine if whistled [s] distortions occur more frequently in certain phonetic contexts than in others. However, the following observations were made. Of the nineteen significant utterances, ten initiated an accented syllable and nine initiated an unaccented syllable. Seventeen of the nineteen significantly judged utterances occurred in blends. Phonetic context may likely be an important factor in therapy of the whistled [s]. In therapy it might be advantageous to "deep-test" the whistled [s] using the words from this study to determine which phonetic contexts are less severely or more severely distorted.

This word list (Appendix I), at least, is a good test to be given before dismissal from therapy. If the client can correctly articulate the /s/ in all phonetic contexts, then he is well on his way toward having corrected his /s/.

Since this study was unsuccessful in providing a recorded severity scale of the whistled [s], such a scale remains to be devised. Perhaps a smaller equal-appearing interval scale could be used successfully, or perhaps the criteria for subjects or judges could be changed to serve the purpose more perfectly.

APPENDIX I

EXPERIMENTAL WORD LIST

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
1.	scene	[sin]	[si]	IA
2.	speech	[spitʃ]	[spi]	IA
3.	scheme	[skim]	[ski]	IA
4.	steamboat	['stim,bot]	[sti]	IA
5.	sleep	[slip]	[sli]	IA
6.	sneak	[snik]	[sni]	IA
7.	sweet	[swit]	[swi]	IA
8.	spree	[spri]	[spri]	IA
9.	spleen	[splin]	[spli]	IA
10.	squeek	[skwik]	[skwi]	IA
11.	screech	[skritʃ]	[skri]	IA
12.	stream	[strim]	[stri]	IA
13.	kerosene	['kɛrə,sin]	[si]	IUA
14.	speedometer	[spi'dɑmətɚ]	[spi]	IUA
15.	city	['sItI]	[sI]	IA
16.	spirit	['spIrIt]	[spI]	IA
17.	skin	[skIn]	[skI]	IA
18.	stick	[stIk]	[stI]	IA

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
19.	slid	[slId]	[slI]	IA
20.	smear	[smIr]	[smI]	IA
21.	sniff	[snIf]	[snI]	IA
22.	switch	[swIt͡ʃ]	[swI]	IA
23.	sprig	[sprIg]	[sprI]	IA
24.	splinter	[splIntɚ]	[splI]	IA
25.	squint	[skwInt]	[skwI]	IA
26.	scripture	['skrIpt͡ʃɚ]	[skrI]	IA
27.	strictly	['strIktlI]	[strI]	IA
28.	bicycle	['baI,sIkl̩]	[sI]	IUA
29.	expidition	[ˌɛkspI'dI͡ʃən]	[spI]	IUA
30.	candlestick	['kændl̩,stIk]	[stI]	IUA
31.	blacksmith	['blæksmIθ]	[smI]	IUA
32.	exquisite	['ɛkskwIzIt]	[skwI]	IUA
33.	manuscript	['mænjə,skrIpt]	[skrI]	IUA
34.	tapestry	['tæpIstrI]	[strI]	IUA
35.	safe	[sef]	[se]	IA
36.	spade	[sped]	[spe]	IA
37.	escape	[ə'skep]	[ske]	IA
38.	state	[stet]	[ste]	IA
39.	slate	[slet]	[sle]	IA
40.	snail	[snel]	[snɛ]	IA
41.	persuade	[pɚ'swed]	[swe]	IA

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
42.	explain	[Ik'splen]	[sple]	IA
43.	exclaim	[Ik'sklem]	[skle]	IA
44.	scrape	[skrep]	[skre]	IA
45.	straight	[stret]	[stre]	IA
46.	wholesale	['hol,sel]	[se]	IUA
47.	landacape	['lænskep]	[ske]	IUA
48.	real estate	['rilə,stet]	[ste]	IUA
49.	several	['sɛvrəl]	[sɛ]	IA
50.	speck	[spɛk]	[spɛ]	IA
51.	schedule	['skɛdʒUl]	[skɛ]	IA
52.	step	[stɛp]	[stɛ]	IA
53.	sledge	[slɛdʒ]	[slɛ]	IA
54.	smelt	[smɛlt]	[smɛ]	IA
55.	snare	[snɛr]	[snɛ]	IA
56.	sweat	[swɛt]	[swɛ]	IA
57.	spread	[spred]	[sprɛ]	IA
58.	splendid	[splɛndId]	[splɛ]	IA
59.	square	[skwɛr]	[skwɛ]	IA
60.	stretch	[strɛtʃ]	[strɛ]	IA
61.	celebration	[,sɛlə'breʃən]	[sɛ]	IUA
62.	expectation	[,ɛkspɛk'teʃən]	[spɛ]	IUA
63.	footstep	['fUt,atɛp]	[stɛ]	IUA
64.	widespread	['waId,sprɛd]	[sprɛ]	IUA

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
65.	sat	[sæt]	[sæ]	IA
66.	span	[spæn]	[spæ]	IA
67.	scandal	['skændl̩]	[skæ]	IA
68.	stack	[stæk]	[stæ]	IA
69.	slap	[slæp]	[slæ]	IA
70.	smash	[smæʃ]	[smæ]	IA
71.	snatch	[snætʃ]	[snæ]	IA
72.	swam	[swæm]	[swæ]	IA
73.	sprang	[spræŋ]	[spræ]	IA
74.	splash	[splæʃ]	[splæ]	IA
75.	scrap	[skræp]	[skræ]	IA
76.	strap	[stræp]	[stræ]	IA
77.	salvation	[sæl'veʃən]	[sæ]	IUA
78.	Scandinavian	[ˌskændəˈnevɪən]	[skæ]	IUA
79.	Constantinople	[ˌkɑnstæntəˈnopl̩]	[stæ]	IUA
80.	extract	['ɛkstrækt]	[stræ]	IUA
81.	sergeant	['sɑrdʒənt]	[sɑ]	IA
82.	sparta	['spɑrtə]	[spɑ]	IA
83.	microscopic	[maɪkrə'skɑpɪk]	[skɑ]	IA
84.	star	[stɑr]	[stɑ]	IA
85.	slav	[slɑv]	[slɑ]	IA
86.	smock	[smɑk]	[smɑ]	IA
87.	snarl	[snɑrl]	[snɑ]	IA
88.	swan	[swɑn]	[swɑ]	IA

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
89.	squadron	['skwɑdrən]	[skwɑ]	IA
90.	astronomer	[ə'strɑnəmɚ]	[strɑ]	IA
91.	sardine	[sɑr'din]	[sɑ]	IUA
92.	spontaneity	[ˌspɑntə'niətI]	[spɑ]	IUA
93.	starvation	[stɑr'veʃən]	[stɑ]	IUA
94.	soldier	['soldʒɚ]	[so]	IA
95.	spoke	[spok]	[spo]	IA
96.	scope	[skop]	[sko]	IA
97.	stone	[ston]	[sto]	IA
98.	slope	[slop]	[slo]	IA
99.	smoke	[smok]	[smo]	IA
100.	snow	[sno]	[sno]	IA
101.	swore	[swor]	[swo]	IA
102.	explore	[Ik'splor]	[splo]	IA
103.	scroll	[skrol]	[skro]	IA
104.	strode	[strod]	[stro]	IA
105.	also	['ɔlso]	[so]	IUA
106.	export	['ɛksport]	[spo]	IUA
107.	telescope	['tɛləˌskop]	[sko]	IUA
108.	limestone	['laImˌston]	[sto]	IUA
109.	soft	[sɔft]	[sɔ]	IA
110.	spawn	[spɔn]	[spɔ]	IA
111.	scald	[skɔld]	[skɔ]	IA

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
112.	stall	[stɔl]	[stɔ]	IA
113.	slaughter	['slɔtə˞]	[slɔ]	IA
114.	small	[smɔl]	[smɔ]	IA
115.	snort	[snɔrt]	[snɔ]	IA
116.	swarm	[swɔrm]	[swɔ]	IA
117.	sprawl	[sprɔl]	[sprɔ]	IA
118.	squall	[skwɔl]	[skwɔ]	IA
119.	scrawl	[skrɔl]	[skrɔ]	IA
120.	strong	[strɔŋ]	[strɔ]	IA
121.	Arkansas	['ɑrkən,sɔ]	[sɔ]	IUA
122.	manslaughter	['mæn,slɔtə˞]	[slɔ]	IUA
123.	suit	[sut]	[su]	IA
124.	spool	[spul]	[spu]	IA
125.	schooner	['skunə˞]	[sku]	IA
126.	stool	[stul]	[stu]	IA
127.	slew	[slu]	[slu]	IA
128.	smoothly	[smuðlI]	[smu]	IA
129.	swoon	[swun]	[swu]	IA
130.	exclude	[Ik'sklud]	[sklu]	IA
131.	screw	[skru]	[skru]	IA
132.	strewn	[strun]	[stru]	IA
133.	superintendent	[,supəˈ In'tɛndənt]	[su]	IUA
134.	tablespoon	['tebl̩,spun]	[spu]	IUA

NO.	WORD	PHONETIC TRANSCRIPTION	PHONETIC CONTEXT	INITIATES ACCENTED (IA) OR INITIATES UNACCENTED (IUA) SYLLABLE
135.	forsook	[fə'sUk]	[sU]	IA
136.	stood	[stUd]	[stU]	IA
137.	superb	[sU'pɜb]	[sU]	IUA
138.	summer	['sʌmə]	[sʌ]	IA
139.	sponge	[spʌndʒ]	[spʌ]	IA
140.	sculpture	['skʌlptə]	[skʌ]	IA
141.	study	['stʌdI]	[stʌ]	IA
142.	slumber	['slʌmbə]	[slʌ]	IA
143.	smother	['smʌðə]	[smʌ]	IA
144.	snuff	[snʌf]	[snʌ]	IA
145.	swung	[swʌŋ]	[swʌ]	IA
146.	sprung	[sprʌŋ]	[sprʌ]	IA
147.	scrub	[skrʌb]	[skrʌ]	IA
148.	strut	[strʌt]	[strʌ]	IA
149.	support	[sə'port]	[sə]	IUA
150.	inspiration	[ˌInspə'reʃən]	[spə]	IUA
151.	constant	['kɑnstənt]	[stə]	IUA
152.	explanation	[ˌɛksplə'neʃən]	[splə]	IUA
153.	exclamation	[ˌɛkskla'meʃən]	[sklə]	IUA
154.	catastrophe	[kə'tæstrəfI]	[strə]	IUA

APPENDIX II

SAMPLE ANSWER SHEET

19-105-also	1	2	3	4	5	6	7	8	9
17-103-scroll	9	8	7	6	5	4	3	2	1
11-140-sculpture	1	2	3	4	5	6	7	8	9
10-25-squint	9	8	7	6	5	4	3	2	1
2-145-swung	1	2	3	4	5	6	7	8	9
19-143-smother	9	8	7	6	5	4	3	2	1
19-129-swoon	1	2	3	4	5	6	7	8	9
14-96-scope	9	8	7	6	5	4	3	2	1
15-35-safe	1	2	3	4	5	6	7	8	9
16-79-Constantinople	9	8	7	6	5	4	3	2	1
11-151-constant	1	2	3	4	5	6	7	8	9
10-15-city	9	8	7	6	5	4	3	2	1
19-144-snuff	1	2	3	4	5	6	7	8	9
12-104-strode	9	8	7	6	5	4	3	2	1

APPENDIX III

INSTRUCTION SHEET

1. Operational definition of a whistled [s]: For the purpose of this study a whistled [s] is defined as an /s/ production which is accompanied by flow of the constricted breath stream across the edges of the central incisors in such a fashion as to impart a near-tonal quality of the resultant sound.

2. The nine-point equal-appearing interval scale is being used. The numbers run from 1 to 9, then 9 to 1, etc. Don't let this confuse you! Be sure you circle the number you intend to be your judgement.

3. Circle one number per line.

4. Number 1 represents a normal [s]. Number 9 represents the most severe whistled [s] you've ever heard. Place the productions of /s/ you hear on the severity scale from 1 to 9.

5. Judge quickly. You may change a judgement, but it is preferred you rely on your first impression.

6. The words are spaced approximately every two seconds.

7. Make sure the word you hear and judge matches the word you see on the line you mark. Use the card provided to help you keep your place.

8. Be careful not to turn more than one page at a time.

9. Listen only for the production of /s/. Do not judge any other sound, any mispronunciations of entire words, or accents incorrectly placed.

10. If you lose your place, miss a word, or must stop the judging

procedure for any reason, do so. You may ask to have a word (or words) repeated when necessary. In such a case, other judges who have already judged the word to be repeated should not change their judgements.

11. We will take a two minute break each fifteen minutes to give you an opportunity to relax and change your position.

12. Are there any questions?

13. The first page is a practice list of ten words which will not be used for analysis. You will now begin the practice judging.

14. Are there any other questions?

15. Begin judging.

LIST OF REFERENCES

Books

1. Ferguson, G. A. Statistical Analysis in Psychology and Education. New York, Toronto, London: McGraw-Hill Book Company, Inc., 1959.

2. Friend, J. H., and Guralnik, D. B. (gen. ed.). Webster's New World Dictionary of the American Language, College Edition. Cleveland and New York: The World Publishing Co., 1956.

3. Guilford, J. P. Psychometric Methods. New York: McGraw Book Co., Inc., 1936.

4. Kenyon, J. S. and Knott, T. A. A Pronouncing Dictionary of American English. Springfield, Mass., G. & C. Merriam Co., Publishers, 1953.

5. Miner, L. Speech Clinician's Handbook. Springfield, Illinois: Charles C. Thomas, Publishers, in press.

6. Thorndike, E. L. The Teacher's Word Book. New York: Teacher's College, Columbia University, 1921.

7. Travis, L. E. Speech Pathology. New York: D. Appleton, 1931.

Journals

1. Cullinan, W., Prather, E., and Williams, D. Comparison of procedures for scaling severity of stuttering. JSHR, 1963, 6, 187.

2. Curry, R. L., Kennedy, L., Wagner, L., and Wilke, W. A phonographic scale for the measurement of defective articulation. JSD, 1943, 8, 123-126.

3. Irwin, R. B., and Krafchick, I. P. An audio-visual test for evaluating the ability to recognize phonetic errors. JSHR, 1965, 8, 281-290.

4. Jordan, E. P. Articulation test measures and listener rating of articulation defectiveness. JSHR, 1960, 3, 303-319.

5. Lewis, D., and Sherman, D. Measuring the severity of stuttering. JSHD, 1951, 16, 326.

6. Morrison, S. Measuring the severity of articulation defectiveness. JSHD, 1955, 20, 347-351.

7. Reid, G. The etiology and nature of functional articulatory defects in elementary school children. JSD, 1947, 12, 143-150.

8. Sherman, D., and Moodie, C. Four psychological scaling methods applied to articulation defectiveness. JSHD, 1957, 22, 698-706.

9. Templin, M. C. Norms on a screening test of articulation for ages three through eight. JSHD, 1953, 18, 323-331.

10. Wood, K. S. Measurement of progress in the correction of articulatory speech defects. JSHD, 1949, 14, 171-174.

11. Wright, H. N. Reliability of evaluations during basic articulation and stimulation testing. JSHD, Monogr. Suppl., 1954, 4, 20-27.