Proceedings of The Institute of Acoustics

PERCEPTUAL AND ACOUSTIC CORRELATES OF STUTTERER'S SPEECH BEFORE AND AFTER TREATMENT IN A PRECISION FLUENCY SHAPING PROGRAM

M.-C. Franken (1) and L. Boves (2)

(1) Dept. of Speech Pathology, Nijmegen University
(2) Institute of Phonetics, Nijmegen University

INTRODUCTION

Recently there has been an increased interest in so called fluency shaping programs for the treatment of stuttering. One prominent example of such a therapy is the Precision Fluency Shaping Program designed by Webster [1]. This kind of therapy attempts to reduce the frequency of dysfluencies by teaching the stutterer to use specially designed phonation, articulation and breathing strategies. In Webster's program these strategies are operationalised in the form of a number of precisely described *target behaviours*. Clients are trained to consciously practice these target behaviours in an intensive course that takes up at least 120 hours. Usually, the complete course is given in three weeks of full time training in the clinic, with the added obligation for the clients to keep consciously practising the target behaviours while they are at home. Sometimes the course is spread over more than three weeks, in which case the contacts between client and therapist are less intensive.

The most important taget behaviours are:

1. *Slow syllable speech.* Each syllable should last precisely for 2 seconds, the first of which should be taken up by the syllable initial sound. The articulation positions of the sounds should be held stable and transitions within a syllable should be extremely gradual and consciously performed.
2. *Full Breath.* Breathing should be conscious, deep, slow and at regular intervals.
3. *Gentle Voice Onset.* This target behaviour is trained with the help of a feedback device that turns on a light when the amplitude envelope of the speech rises too steeply.

A complete therapy program divides into four successive phases. During the first phase all three target behaviours are learned in the way as described above. In this phase the behaviours are practised consciously in extremely slow and paced speech. In the second phase the target behaviours have to be stabilised and automatised, while the speech tempo is slightly increased. During the third phase the speech tempo is increased still further, until a result is reached that that might be characterised as 'slow-normal' speech. The last phase consist of the transfer of the newly acquired speech behaviour to situations outside the clinic and the family.

From the description of Webster's PFSP given above it should be obvious that the therapy affects the speech of the stutterers in many more ways besides from (hopefully) reducing the number of dysfluencies. It is not at all clear whether all side effects are desirable. This is one of the reasons why a formal evaluation of the therapy is needed, which is based on a comprehensive description of the speech before and after treatment. If nothing else, such an evaluation might point out some aspects where the therapy program could be improved or completed.

Comprehensive descriptions of the quality of a speech sample are difficult to make. Also, they tend to be extremely time consuming and therefore expensive.

In this paper we will present and compare three different approaches, viz. ratings of a group of essentially untrained raters who judge speech samples on a set of 15 seven-point scales, expert ratings of a number of features that are characteristic of the speech of stutterers, and a number of automatic acoustic measurements.

SPEECH MATERIAL

Our work is based on speech samples from 21 clients, 17 adult males, 3 adult females and one male adolescent who went through a Dutch adaptation of Webster's Precision Fluency Shaping Program. Each client was recorded on three occasions, viz. on the first day of the therapy, on the last day of the therapy and about six months after the end of therapy. Use was made of spontaneous speech, produced while the clients gave their opinion and comments on a newspaper article. Approximately 45 seconds of speech was taken from each recoding session, omitting the first 30 seconds of the recordings. To this material recordings were added from five stutterers who were not treated and from five non-stuttering control speakers. The ten speakers last mentioned were recorded twice under the same conditions as the treated stutterers. This made for a total of 83 speech samples.

Table 1. Summary of the rating scales and their reliabiliy

Pleasant <--> Unpleasant	.98
Slow <--> Quick	.99
Low Pitch <--> High Pitch	.96
Soft <--> Loud	.93
Husky <--> Not Husky	.72
Broad <--> Cultured	.97
Expressionless <--> Expressive	.97
Monotonous <--> Melodious	.97
Slovenly <--> Polished	.95
Dull <--> Clear	.95
Weak <--> Powerful	.95
Shrill <--> Deep	.95
Dragging <--> Brisk	.98
Ugly <--> Beautiful	.97
Unnatural <--> Natural	.97

PERCEPTUAL RATINGS BY UNTRAINED JUDGES

For the perceptual rating the 83 stimuli were put in a two different random orders. Each tape was rated by a different group of listeners. Both groups consisted of first year students of logopedics, who may be considered as essentially naive listeners at the time of the experiment. One group consisted of 29 members, the other one of 17. The ratings of the two groups were combined to a single set of scores. The 15 bipolar scales used in the rating experiment are shown in Table 1. In that table the effective reliability of the scores is also given. Except for the scale Unnatural <--> Natural the scales are taken from [2].

Factor analysis of the scores on the scales resulted in a 5-factor solution.

The first factor was identified as a combined tempo and general evaluation factor. It accounts for 64% of the variance explained, which shows that tempo is the single most important aspect to characterise speech of stutterers. The second factor pertains to ariculation quality, the third to voice dynamics, the fourth to voice pitch, and the fifth to potency.
Factor scores were computed for the 83 stimuli and the 63 scores pertaining to the speech of the treated stutterers were then subjected to an analysis of variance, using a single factor design with repeated measurements in three conditions (pretherapy, post therapy, and follow up). The aim of this analysis was to obtain insight into the effects of the therapy, both intentional and unintentional.
The average factor scores on the first (tempo/evaluation) factor appeared to drop slightly in the post therapy condition. Six month after therapy, however, the average score in this factor is well above the remaning two conditions. Thus it appears that it is only six month after the completion of the therapy that clients' speech is rated more positively. Just after therapy the judgments tend to be even more negative than before therapy started. A complete account of the results of the factor analysis and the analysis of variance can be found in [3].

EXPERT RATINGS

Table 2 defines the features of the speech that were established by one expert clinician. These features were counted or rated for all speech samples, including those of the non-stuttering controls. The counting of the number of linguistic syllables is based on a transcript of the texts. In making the transcripts exact repetitions of complete words were omitted. The number of times such repetitions ocur is counted and used as an individual measure. The conversion to number of syllables per unit speech time is effected by measuring the time during which speech sound is present. This is done in a fairly crude way, viz. by holding an electronic stopwatch by hand during the audible speech pauses.
An attempt has been made to distinguish between repetitions of individual sounds and repetitions of syllable-sized units. Vowel repetitions have been counted as syllable repetitions only if standard syllabification rules identify the vowel as a complete syllable. CV repetitions have always been considered as syllable repetitions, even if according to standard syllabification rules the syllable structure is CVC.
The measures 2., 3., and 4. only count the number of syllables on which some kind of dysfluency is present; these measures do not reflect the number of dysfluencies present on each of the syllables. Counting the exact number of times a syllable or sound is repeated, however, is extremely difficult. At the same time it is far from obvious how these numbers should be interpreted.
Dysfluencies were rated as blockades if a sound was terminated in an excessively abrupt way, while at the same time giving the rater the impression of excessive laryngeal tension with the speaker. A blockade was also indicated if a speech pause was inserted at a non-syntactically motivated point or if a syntactically normal pause lasted too long, while again the auditory impression of the sounds following the pause was one of excessive muscular tension. Interjections specific for stuttering are also characterised by the auditory impression of very high articulatory tension.
The number of syllables containing stuttering dysfluencies is established by

subtracting the number of syllables in complete word repetitions from the number of syllables containing at least one dysfluency.

Table 2. Definition of the features for use by experts

1.	Number of linguistic syllables in the sample per uinit speech time
2.	Number of syllables containing at least one dysfluency
3.	Number of sound repetitions not to be considered as syllable repetitions
4.	Number of V- or CV repetitions, considered as syllable repetitions
5.	Number of prolongations of syllable initial sound
6.	Number of prolongations of syllable internal/final sounds
7.	Number of blockades
8.	Number of interjections specific for stuttering
9.	Number of syllables containing stuttering dysfluencies
10.	Number of syllable repetitions that are part of word repetitions
11.	Number of filled pauses
12.	Rhythm (non-scoreable, normal, slightly/fairly/strongly deviant)
13.	Voice dynamics (same scoring alternatives as with rhythm)
14.	Breathing (same scoring alternatives as with rhythm)
15.	Laryngeal articulation (same alternatives as with rhythm)
16.	Supra-laryngeal articulation (same alternatives as with rhythm)

The remaining features in Table 2, including the five that were rated on a four point scale, are self-explaining. It was decided to add a fifth alternative to the rating scales in order to be able to cope with speech samples that were so abnormal that valid and sensible ratings of a feature was impossible. This alternative was chosen for a small number of stimuli from the pre therapy condition when rating the rhythm and voice dynamics scales.
Analyses of variance of the individual measures for the treated stutterers showed significant condition effects for all but three features, viz. 5., 6., and 7. The measures that did discrimunate between the conditions at the .01 level of significance all showed a V- or inverted V shape, i.e., the post therapy condition is very different from the remaining conditions. In the post therapy condition the number of dysfluencies and the degree of deviant breathing, laryngeal and supra-laryngeal behaviour are considerably lower than in the pre therapy and in the follow up condition. For most clients, the numbers in the follow up condition are still appreciably below the pre therapy condition. Thus it appears that in general the results of the therapy are retained over a period ofsix months. With respect to the rhythm and voice dynamics measures, however, the post therapy condition is significantly more deviant than the pre therapy and follow up conditions. Thus we see that in the follow up condition the number of dysfluencies which are typical of stuttering is increased relative to the situation just after therapy, but that this increase of dysfluencies is compensated by an increased normalcy of the suprasegmental aspects of the speech.

ACOUSTIC MEASUREMENTS

The acoustic analyses carried out on the speech samples are listed in Table 3. Most of the measures are self-evident and have been described and used previously [2,4]. The numbers 6 to 9 describe the global shape of the long term

Table 3. Definition of the acoustic measures

1. Articulation rate: Syllables/second speech, exclusive of pauses
2. F0 mean
3. F0 standard deviation
4. Pitch perturbation
5. Proportion of the speech which is periodic (voiced)
6. Spectral level in the region of the first formant
7. Spectral slope below first formant
8. Spectral slope towards 1.5 kHz
9. Spectral slope towards 5.0 kHz
10. Neutral Articulation Setting
11. Standard deviation of first formant
12. Average intensity
13. Intensity variation

average critical band spectrum of the voiced parts of the speech. 'Neutral articulation setting' is defined as the Euclidian distance of the average frequencies of the first four formants of the speaker to the hypothetical neutral point of 500, 1500, 2500, and 3500 Hz in a four-dimensional space.
Analyses of variance were carried out in order to identify those measures that discriminate between the three experimental conditions for the group of treated stutterers. Significant effects were only found for articulation rate, pitch perturbation, F0, F0 variation, proportion of voiced segments and intensity variation. The results of the analysis of variance are summarized in Table 4. Articulation rate in the post therapy condition does not differ form the pre therapy value. In the follow up condition, however, articulation rate appears to have improved. The higher proportion of voiced segments in the post therapy condition is a direct consequence of the remainders of the target behaviours of slow syllable speech and gentle voice onset. The same is probably true for the decrease in F0 variation and intensity variation. The significant drop in average F0 in the post therapy and follow up conditions is not easy to explain. One possibility is that this drop is due to a decrease in overall muscular tension as a result of the therapy. If that hypothesis holds, one would expect a corresponding effect in some of the spectral measures, which is, however, not present. An alternative explanation might be that the drop in average F0 is the result of the lowered level of F0 variation; this would imply that most of the pitch movements in a more lively type of speech serve to raise the voice pitch.

Table 4. Summary of results of analysis of variance on acoustic measures

Measure	Overall	Condition Means				Post Hoc Contrasts		
	Signif.	Pre	Post	Follow		1-2	1-3	2-3
Rate	.004	2.11	1.95	2.68	syl/sec	ns	.003	.001
F0 Mean	.02	126.9	120.0	121.5	Hz	.025	.016	ns
F0 variation	.005	20.9	13.9	16.7	Hz	.008	.014	ns
Pitch Perturb.	.01	.38	.50	.44		.007	ns	ns
Prop. Voiced	.001	42.2	67.4	58.0	%	.001	.003	.001
Intens. Var.	.001	9.57	6.81	8.14	dB	.001	.027	.008

RELATIONS BETWEEN THE THREE DESCRIPTIONS

Having three different descriptions of the same set of speech samples it is interesting to see to what degree they coincide. One way of carrying out such a study is by means of multiple regression analyses. Because we feel that for the time being perceptual measures are more meaningful for the description of speech quality than acoustic measures, we have restricted the regression study to attempts to predict perceptual scores by acoustic measures. For the study of the relation between acoustic measures and the ratings of the naive judges the scores of the individual stimuli on the five factors described above have served as the point of departure. This enables us to present the most salient results in a compact manner.

The relation between ratings of naive judges and acoustic measures

The tempo - general evaluation factor. Over 80% of the variance in the scores on this factor is accounted for by articulation rate. This result is not very surprising. Articulation rate is the acoustic measure that best captures the most salient and prominent feature that makes the speech samples differ between each other.

The articulation quality factor. Only 30% of the variance in the factor scores can be accounted for by a combination of the average formant position in the F1 - F4 space and the spectral slope below the F1 region. Previous attempts to explain ratings of articulation quality by means of acoustic measurements did nor fare much better. Either do we need much more sophisticated acoustic measures of articulation quality, or what comes out of the perceptual rating experiments does not describe articulation quality proper, but rather represents some complex overall evaluation of speech quality, in which articulation is only one of a multitude of relevant aspects.

The voice dynamics factor. Although one would expect this factor to be related to acoustic measures like F0 variation and intensity variation, the only acoustic measure that explains an appreciable amount of variance (viz. 20%) in the factor scores appears to be articulation rate. In the material under study, where rate is the aspect that varies most widely, more plausible features like pitch and loudness variation hardly seem to influence the impression that untrained listeners obtain from the speech.

The pitch factor. Half of the variance in the scores on the pitch factor is accounted for by average F0. Addition of some spectral measures, particularly the location of the average formant frequencies in the F1-F4 space, raises the percentage of explained variance to over 60%. This result is in accordance with previous findings indicating that in the rating of pitch level by untrained judges spectral properties of the signal play a role besides fundamental frequency [5].

The potency factor. Only 33% of the total variance is explained by the combination of articulation rate and intensity variation. The difficulty with this factor is not due to its slightly esoteric character. The factor is entirely made up by the scales 'weak <--> strong' and 'soft <--> loud'. Of these, the less physical one, viz. 'weak <--> strong' fares much better (with 66% of its variance explained) than the scale 'soft <--> loud' (38% explained).

Relations between acoustic measures and expert ratings

There appeared to be little relation between the expert ratings of speech behaviour that is typical for stutterers on the one hand and global acoustic measures on the other. For most of the feautures rated by the expert this finding did not come as a surprise. Given the characteristics of our acoustic measures one may hardly expect that they are able to predict the number of dysfluencies in a sample of speech. The only exception to the rule that the proportion of variance in the expert ratings that can be explained by the acoustic measures never exceeds 20% are the ratings of deviances in speech rhythm and dynamics. These features appear to be related to the acoustic measure of speech rate. Overall speech rate accounts for 40% of the variance in the ratings on the rhythm scale and for 34% of the variance in the ratings on the speech dynamics scale.

From these findings one may draw the conclusion that the features traditionally used to characterize the speech behaviour of stutterers are almost completely independent from what can be obtained from automated acoustic measures. The interpretation that must be given to this conclusion is, however, far from evident. In fact, two completely contradictory interpretations seem equally probable. One might say that the result shows that acoustic measurements have nothing to contribute if it comes to diagnosing or evaluating the speech of clients with a stuttering problem. Others might favour the opposite interpretation, viz. that acoustic measures, once we have learned how to use them, may give the therapist access to a number of features of the stutterer's speech that are not captured in the traditional descriptions.

Relations between the ratings of naive listeners and expert ratings

The large difference between the type of features rated by the untrained listeners and the features addressed in the expert ratings leads one to predict that there will be no direct and simple relations between the two sets of scores. In a first analysis we have restricted our attention to the expert ratings based on counting dysfluencies (the first 11 features in Table 2). We have carried out two sets of analyses, once with the expert ratings as the criteria and another time with the expert ratings as predictors of the scores of the untrained listeners.

From the results of the analyses it appeared that the expert ratings cannot be predicted by the scores of the naive listeners. The total amount of variance in the expert ratings that can be explained by the ratings of the naive listeners does not exceed 27% for any of the rating scales.

Employing the expert ratings as predictors of the scores of the untrained listeners leads to results that are more promising and insight lending. 40% of the variance in the scale values on the scale 'slow <--> quick' is accounted for by a combination of the number of linguistic syllables in the stimulus per unit time of speech and the number of syllables containing a stuttering dysfluency. The same pair of features accounts for 35% of the variance in the ratings on the scale 'dragging <--> brisk', for 39% of the variance in the scale values on 'expressionless <--> expressive', and for 42% of the variance on the scale 'monotonous <--> melodious'.

Because we had the impression that the expert ratings should also capture many of the aspects of the speech that determined the global evaluation scores of the naive listeners, we looked for transformations that can be applied to the expert ratings and that improve their correlation with the scores of the naive listeners. One possibility is to compute the total number of 'core stutters', i.e., the sum of the expert's counting of the dysfluency types 3.-8. inclusive. This measure did not account very well for the scores of the naive lsiteners either. Separate analyses of the pre, post and follow up conditions showed that the number of core stutters does indeed account for an appreciable amount of the variance in the scores on scales like 'ugly <--> beautiful', but only in the pre therapy and follow up condition. In the post therapy condition the measure does not work at all. Obviously, in the post therapy condition, where the number of core stutters is very small and where the quality of the speech is completely determined by extremely low rate and voice dynamics, other features than the proportion of dysfluencies are decisive. Thus it appears that a valid evaluation of the speech quality after a PFSP should account for other aspects besides the number of dysfluencies. A compromise solution on which we are presently working might be to weigh the number of dysfluencies in some way by speech rate and voice dynamics.
Such a weighing should not simply consist of a linear combination. Inclusion of the expert ratings of rhythm, voice dynamics, breathing, laryngeal and supra-laryngeal behaviour did not improve the situation very much. Only the ratings of the naive listeners on the scales 'monotonous <--> melodious' and 'expressionless <--> expressive' can be explained by the expert ratings of deviant rhythm (46% of the vaiance is accounted for).

CONCLUSIONS

From our work it first appears that the PFSP as it is applied in Holland does indeed help in reducing the number of stuttering dysfluencies. Particularly in the post therapy measurements, however, a number of undesirable side effects are present in the speech of the clients. These side effects may remain completely unnoticed if a description of the speech is restricted to a counting of the number of dysfluencies per unit time, or even per syllable.
The second conclusion that can be drawn is that different approaches to the description of the speech quality lead to different results. Most of these differences seem to be due to the fact that the individual approaches address different aspects of the speech.

REFERENCES

[1] R.L. Webster, 'Evolution of a target-based behavioral therapy for stuttering', J. Fluency Disorder, Vol. 5, 303-320, (1980).

[2] L. Boves, 'The phonetic basis of perceptual ratings of running speech', Foris Publ. (1984).

[3] M.-C. Franken, 'Perceptual evaluation of a stuttering therapy', in H. Peters & W. Hulstijn (Eds.) Speech Motor Behavior in Stuttering, Springer Verlag, (1986).

[4] R. van Bezooijen, 'Characteristics and recognizability of vocal expressions of emotion', Foris Publ., (1984)