THE RELATION BETWEEN PERCEPTUAL AND CLINICAL PARAMETERS OF VOICE QUALITY OF PATIENTS WITH EARLY GLOTTIC CANCER BEFORE AND AFTER RADIOTHERAPY AND OF NORMAL SPEAKERS

Irma M. de Leeuw

1. INTRODUCTION

From a theoretical point of view it seems trivial to state that there exists a relation between perception of voice aspects and physically measurable parameters: what you can hear, you should be able to measure and vice versa: what you do measure, you should be able to hear.
In practice this assumption is not supported by research: judgements based on auditory perception can be related to acoustical voice parameters only in a difficult and incomplete way. Still, in diagnosis and therapy one starts from judgements based on auditory perception of voice.
The aim of this study is to find a relationschip between perceptual, acoustical and clinical parameters of voices of patients with early glottic cancer who are getting radiotherapy and of the voices of healthy speakers.

This study is part of a co-operation between the Netherlands Cancer Institute (Antoni van Leeuwenhoek Hospital), the Academic Hospital of the Free University of Amsterdam and the Institute of Phonetic Sciences of the University of Amsterdam. In this co-operation a study will be carried out at the Netherlands Cancer Institute by radiotherapist G. Baris on dose response in radiotherapy of early glottic cancer. The purpose of that study is to determine the optimal radiation dose to be delivered to small glottic tumours. Optimal radiation dose should be based upon dose response curves for tumour control and normal tissue complications. One of these complications causes decrease of voice quality.
Voice quality can be measured by perceptual, acoustical, and clinical parameters. Common methods for measuring voice quality in phonetic research are subjective judgements of listeners (perceptual parameters) and acoustical analysis of voice (acoustical parameters). Clinical methods are tests related to aerodynamic efficiency, to fundamental frequency and sound pressure level, to vocal fold vibration, etc. (Hirano, 1990).
From previous research it appears that relations between perceptual and acoustical parameters have been studied more often than relations between perceptual and clinical parameters.

In this paper perceptual and clinical parameters are correlated. This experiment is a pilot study for further research. The chosen clinical parameters are based upon general findings in literature. The results show that correlations between perceptual and clinical parameters exist, although these univariate correlations are not high.The purpose is to find out if there are resemblances between the results from the data which are now available and the results found in literature. In future research, other (acoustical) parameters will be included; also, more extensive statistical methods will be used.

2. EXPERIMENT

The experiment consists of an evaluative rating by 25 untrained and by 8 trained listeners based on text read aloud by 3 groups of speakers. Furthermore, clinical measurements are gathered: phonation quotient, phonetogram and, fundamental frequency.

2.1. Speakers

The speakers in this experiment are divided into three groups. Group 1 and 2 consists of the same 5 patients with early unilateral glottic cancer with no impaired cord mobility. Group 1 is the group of patients before radiotherapy. Group 2 is the group of patients 6 months after radiotherapy. Group 3 is the control group of speakers without early glottic cancer.
The matching between group 1 and 3 includes the following parameters: sex, age, smoking habits, and vocal use (profession/hobby). The matching did take place before radiotherapy. A review of the matching criteria is given in table 1.

Table 1. Matching criteria of patients (speakers 1-5) and control-speakers (speakers 13-18).
(smoking: number of cigarettes per day, according to the speaker;
before radiotherapy -> after radiotherapy)

speaker	sex	age	smoking	smoking	age	sex	speaker
1	m	64	25 ->25	25	64	m	13
3	m	62	12 ->0	10	61	m	15
4	m	71	10 ->0	10	67	m	16
5	m	50	5 ->0	15	44	m	17
6	m	60	25 ->2	30	58	m	18

2.2. Perceptual parameters

The perceptual parameters are judgements of listeners on a test containing semantical bipolar seven point scales. This test is developed to obtain a reliable perceptual description of voice and pronunciation. In various experiments (Fagel & Van Herpt 1983; Van Herpt, 1986) it appears that the chosen 14 scales can be described in a 5-dimensional perceptual space: Voice Appreciation, Articulation Quality, Voice Quality, Pitch, and Tempo. The first dimension, Voice Appreciation, possibly can be broken down into Melodiousness and Evaluation. Dimension 3 can be divided in Clarity and Subjective Strength. Since the topic in this experiment is the quality of voice of pathological voices one additional scale is added: *intelligible-not intelligible.* The semantic scales and the dimensions are given in table 2.

The speakers read aloud a text of about 5 minutes. All texts were recorded using a Philips D6920 MK2 cassette recorder and a Philips N 8214 microphone. Fragments of all texts (ca 1 min.) were copied in random order to one reel tape and this tape was copied to 35 cassette tapes by using a Tandberg TCD 310 cassette recorder and a Revox A 77 tape recorder. The tapes were presented binaurally via a cassette recorder and headphones to a group of 25 students (speech therapy) and 8 speech therapists. The students are considered to be untrained listeners and the speech therapists to be trained listeners.

Results of perceptual parameters

Means and standard deviations are computed for both groups of listeners and for each of the 3 groups of speakers, for scale 1 up to scale 15, see figure 1.

Table 2. Semantic scales

Scale nr.	Scale	Dimension
1	monotonous -- melodious	I Voice Appreciation
2	expressionless -- expressive	A. Melodiousness
3	unpleasant -- pleasant	B. Evaluation
4	ugly -- beautiful	
5	slovenly -- polished	II Articulation Appreciation
6	broad -- cultured	
7	husky -- not husky	III Voice Quality
8	dull -- clear	A. Clarity
9	soft -- loud	B. Subjective Strength
10	weak -- powerful	
11	high -- low	IV Pitch
12	shrill -- deep	
13	slow -- quick	V Tempo
14	dragging -- brisk	
15	not intelligible -- intelligible	

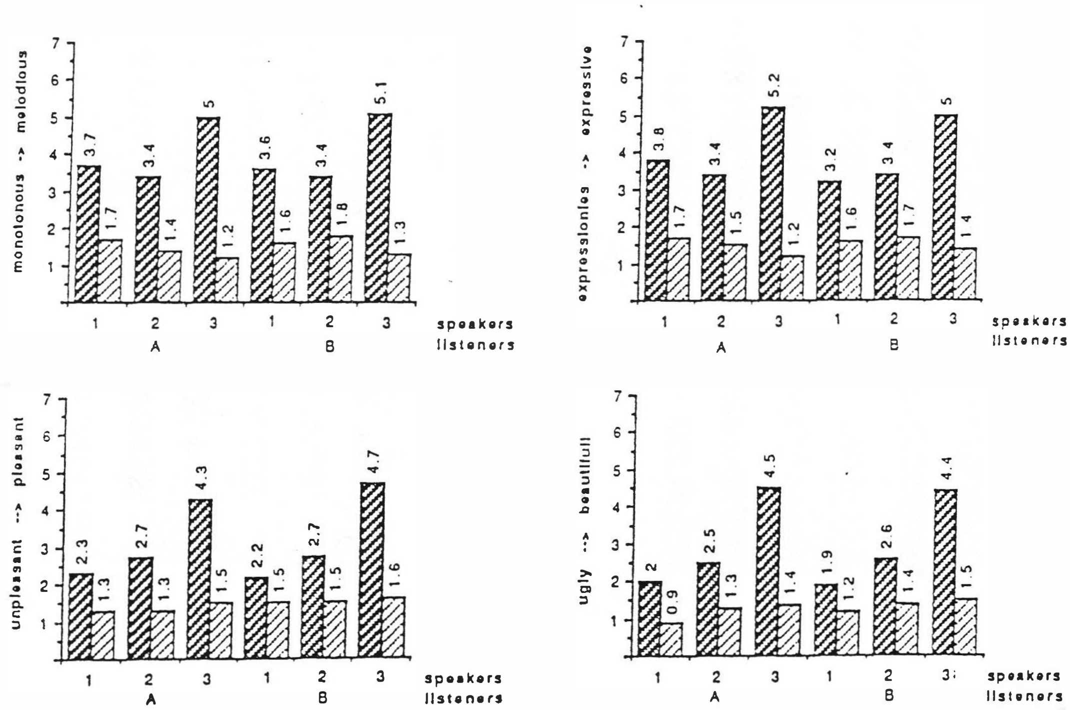

Figure 1 . Means (dark bars) and standard deviations (light bars) per group of listeners per group of speakers for the first 4 scales.

Patients before radiotherapy = group 1 speech therapists = group A
Patients after radiotherapy = group 2 students = group B
Control speakers = group 3

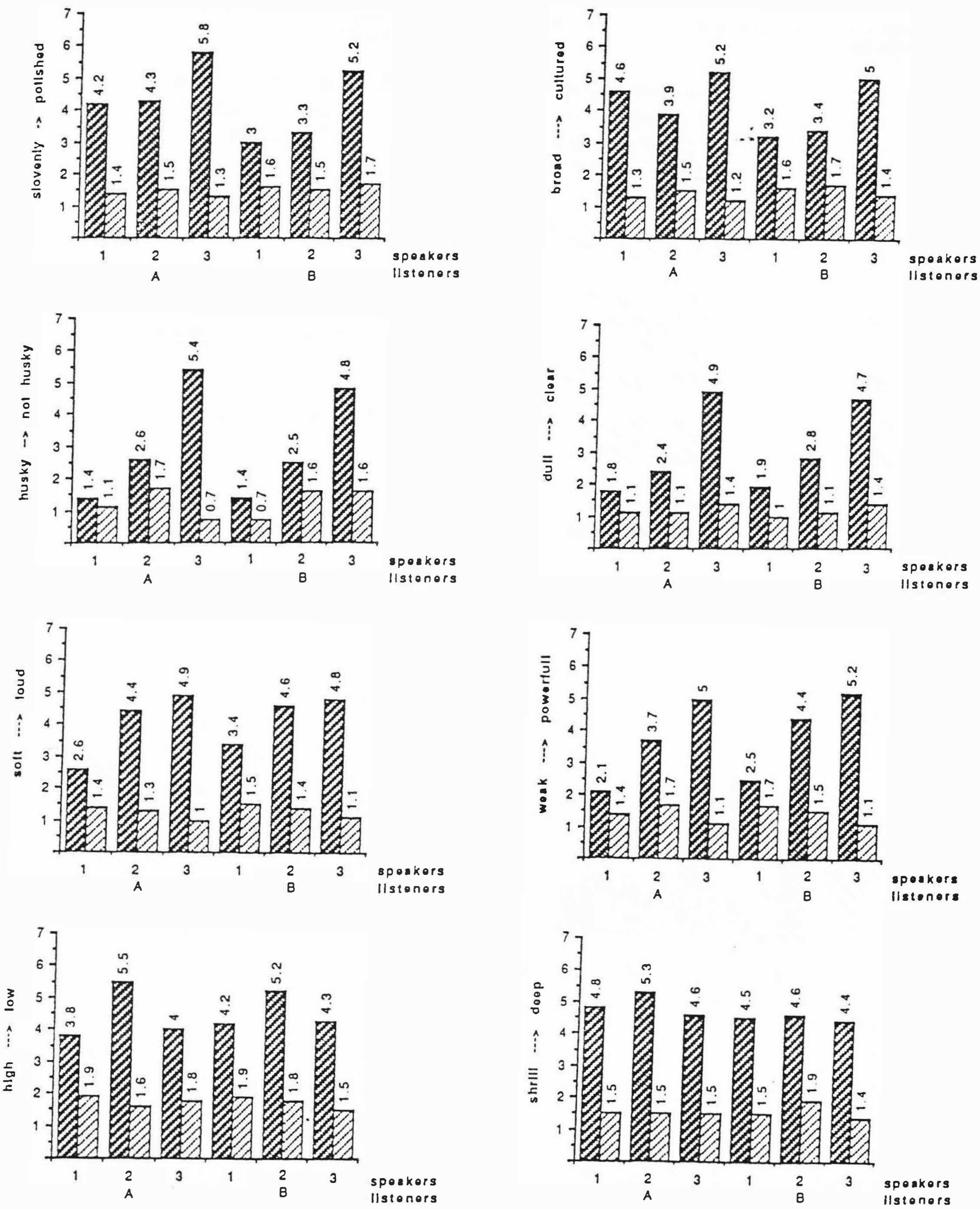

Figure 1 (cont.). Means (dark bars) and standard deviations (light bars) per group of listeners per group of speakers for the next 8 scales.

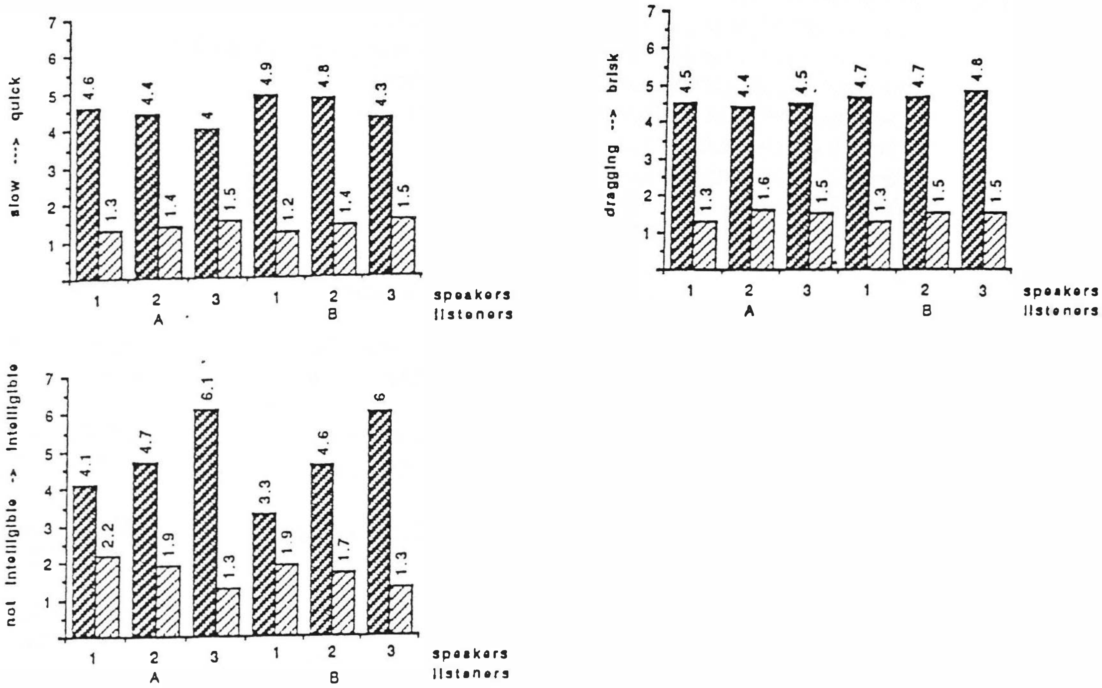

Figure 1 (cont.). Means (dark bars) and standard deviations (light bars) per group of listeners per group of speakers for the last 3 scales.

2.3. Clinical parameters

2.3.1. Phonetogram

The phonetogram is made to determine the frequency and intensity range of phonation. The frequency range of phonation is the difference between the highest and the lowest tone a person can repeat. The intensity range of phonation is the difference between the loudest and the weakest tone a person can repeat at different frequencies. The data are written in the phonetogram.

An experienced speech therapist made the phonetograms. The phonetograph stood in an anechoic room. The speech therapist gave a tone and the speaker repeated it into a microphone. The distance between each speaker and the microphone was the same (30 cm). The speech therapist read the minimal and maximal intensity from the decibel meter and wrote it down in the phonetogram. In this way the whole frequency range of the speaker was examined.

The phonetograms were quantified in the following way:

- the frequency range of phonation was determined in semitones;
- the intensity range of phonation was determined according to Buekers (1980):

  The speaking voice, the middle voice, and the falsetto voice were determined. The speaking voice is the frequency that lies 5 semitones above the lowest tone; the middle voice lies an octave above the speaking voice; the falsetto voice lies an octave above the middle voice.

  The intensity range (difference between the loudest and weakest tone) of the three different voices is determined.

Results of phonetogram

Table 5. Results phonetogram
Lowest and highest tones in Hz and semitones (s.t.).
Speaking voice, middle voice, and falsetto voice in Hz and in semitones; minimal and maximal intensity in dB (min., max).
Frequency range in semitones; intensity range in dB.
Patients before radiotherapy = group 1
Patienten after radiotherapy = group 2
Control speakers = group 3
* no measurements possible

		group 1	group 2	group 3
lowest tone	Hz/s.t.	102.8 / 31,6	78,6 / 27,2	91,6 / 29,4
highest tone	Hz/s.t.	345 / 52	377 / 53,4	412 / 56
speaking voice	Hz/s.t.	139.2 / 36,4	110,8 / 33	122,6 / 34,4
min. / max.	dB/dB	58 / 71	55 / 73	60 / 78,2
middle voice	Hz/s.t.	283.6 / 49,5	215,5 / 44,5	219,2 / 44,7
min. / max.	dB/dB	63,6 / 81,6	64,5 / 80	61,2 / 87
falsetto voice	Hz/s.t.	* / *	409,6 / 55,6	394,5 / 55
min. / max.	dB/dB	* / *	70 / 83,3	73 / 98
frequency range	s.t.	20,4	26.2	26,8
intensity range:				
speaking voice	dB	13,6	18.6	18.3
middle voice	dB	17,6	15.5	26.3
falsetto voice	dB	*	13.3	25

2.3.2. Speaking fundamental frequency

The speaking fundamental frequency was measured in Hz and in semitones by using an electroglottograph (Stopler Teltec GFA06). Electrodes are placed on the skin right and left from the larynx. A high-frequency electrical current passes through the larynx from one electrode to the other. The electrical resistance across the larynx varies with the opening and closing of the glottis. Maximal resistance indicates an open glottis, while minimal resistance indicates vocal fold closure.

Results of speaking fundamental frequency

The results are given in table 3.

Table 3. Means and standard deviations of speaking fundamental frequency in Hz and semitones (s.t.) per group of speakers
Patients before radiotherapy = group 1
Patients after radiotherapy = group 2
Control speakers = group 3

	group 1		group 2		group 3	
	$\bar{x}$	s.d.	$\bar{x}$	s.d	$\bar{x}$	s.d
Hz	155	52,8	107	19,5	120	21
s.t.	38	5,7	32,4	3,3	34,4	2.8

2.3.2. Phonation quotient

The phonation quotient is determined to describe the relation between phonation and respiration. The phonation quotient is defined as the vital capacity divided by the maximum phonation time. The vital capacity was measured by using a spirometer (Pneumoscreen 11/1,84). The speaker took a deep breath, then he did breath into the mouthpiece of the spirometer with closed nose. The maximum phonation time was determined by the time a speaker could sustain an /a/.
The normal mean phonation quotient lies between 120 and 190 ml/s. The upper limit varies from 200 - 300 ml/s (Hirano, 1981). Values above this limit indicate incomplete vocal fold closure. The speaker has an insufficient phonation.

Results of phonation quotient

The results are given in table 4.

Table 4 Means and standard devations, and ranges of the phonation quotient (PQ) in ml/s.
patients before radiotherapy = group 1
patients after radiotherapy = group 2
control speakers = group 3

	group 1			group 2			group 3		
	$\bar{x}$	s.d.	range	$\bar{x}$	s.d.	range	$\bar{x}$	s.d.	range
PQ	385	123	290-590	305	116	200-490	200	43	140-250

4. CORRELATIONS BETWEEN PERCEPTUAL AND CLINICAL PARAMETERS

The chosen parameters in this experiment are based upon general findings in literature. It seems that pitch correlates with fundamental frequency (Kuwabara & Ohgushi, 1984; Eskenazi et al., 1990; Moran & Gilbert, 1984), with frequencies of the formants (Kuwabara & Ohgushi, 1984), and with the amount of noise in the spectrum (Eskenazi et al., 1990; Moran & Gilbert, 1984).
Noise measures also correlate with the voice quality characteristics hoarseness (Murry et al., 1977) and roughness (Hanson & Emanuel, 1979).
Fundamental frequency measures are related to the voice quality characteristics coarse-light (Fritzell & Hammarberg, 1977; Hammarberg et al., 1980) and to roughness, breathiness, and hoarseness (Eskenazi et al., 1990), as well as to chest-head register (Hammarberg et al., 1980).
Formant frequencies can be related to the voice quality characteristics breathiness and nasality (Murry et al., 1977)).
Furthermore voice quality correlates with the slope of the long-term-average-spectrum (LTAS) and with SPL-values in the LTAS (Fritzell & Hammarberg, 1977; Hammarberg et al., 1980), as well as with airflow values (Murry et al.,1977).

If these results are "translated" to the data in the present experiment, correlations between the following variables might be expected:

Dimension III: Voice Quality	< -- >	speaking fundamental frequency
	< -- >	frequency data from the phonetogram
	< -- >	intensity data from the phonetogram
	< -- >	phonation quotient
Dimension IV: Pitch	< -- >	speaking fundamental frequency
	< -- >	frequency data from the phonetogram

4.1. Assumptions

In this pilot study we only verified those relations that are indicated in literature. This means that the following clinical parameters are selected: phonation quotient, speaking fundamental frequency, the lowest, and highest tone from the phonetogram, the speaking voice from the phonetogram, and the loudest and weakest tone of the speaking voice from the phonetogram. Perceptual parameters are the 15 semantical scales.
The semantical scales are considered as interval scales. This means that the distances between the different scale points are considered to be equal. This supposition is based on previous investigations (Blom and Van Herpt, 1976). Means and standard deviations are calculated.
The correlation between the perceptual and clinical parameters is supposed to be linear. The product moment correlation coefficient (Pearson) is calculated for mean judgements per speaker per group of listeners.
The listeners are distinguished into two group: group A consists of 25 students (speech therapy); group B consists of 8 speech therapists. Group A is considered to be untrained, while group B is considered to be trained.

4.2. Results

In table 6 the correlation coefficients are given for the 15 semantical scales with the 7 selected clinical parameters. The coefficients are given only if their absolute value is higher than .45.

Table 6: Correlation coefficients of scale 1 to 15 with clinical parameters.
Only the correlations with an absolute value higher than .45 are indicated. Correlations from students are written on the lefthand side; the correlations from the speech therapists on the righthand side. PQ: phonation quotient; F0: mean fundamental frequency; speak: speaking voice; high: highest tone; low: lowest tone; min: minimal intensity of the speaking voice; max: maximal intensity of the speaking voice.

	clinical parameters						
	PQ	F0	speak	high	low	min	max
Dimensions/scales							
Voice Appreciation							
monotonous-melodious	-.48/						
expressionless- expres.	-.50/-.48						
unpleasant-pleasant	-.45/						
ugly-beautiful							.45/
Articulation Appreciation							
slovenly-polished	-.45/-.47						
broad-cultured							
Voice Quality							
dull-clear	-.50/						.49/.46
husky-not husky							/.52
soft-loud							
weak-powerful	-.48/						
Pitch							
high-low		-.53/-.60	/-.49		-.48/-.56		
shrill-deep							
Tempo							
slow-quick							
dragging-brisk							
not intell.-intell.							

The results show that the expectations based on literature partly come true: the dimension Pitch (the scale *high-low*) correlates with mean fundamental frequency, with the lowest tone (students and speech therapists), and with the speaking voice (speech therapists).
The dimension Voice Quality correlates with maximal intensity of the speaking voice: the scales *dull-clear* (students and speech therapists) and *husky-not husky* (speech therapists). The scales *dull-clear* and *weak-powerful* correlate also with the phonation quotient (students).
The other dimensions seem to correlate with some clinical parameters too. The dimension Voice Appreciation is related to phonation quotient: the scales *monotonous-melodious* (students) and *expressionless-expressive* (students and speech therapists). The scale *unpleasant-pleasant* correlates with phonation quotient (students) and the scale *ugly-beautiful* with maximal intensity of the speaking voice (students).
The dimension Articulation Appreciation (the scale *slovenly-polished*) is related to the phonation quotient (students and speech therapists).
The dimension Tempo and the scale *not intelligible-intelligible* are not related to any of the clinical parameters.

To get a better idea of the various correlations some scatter diagrams are examined. The diagrams of the correlations between the scale *dull-clear* with the phonation quotient and the maximal intensity, and between the scale *high-low* with mean fundamental frequency, speaking voice, and lowest tone are presented here. These scales correlate with more than one clinical parameter; besides, these correlation coefficients appear to be the highest. Judgements from students as well as from speech therapists are examined. The diagrams are given in figure 2a and 2b, respectively.

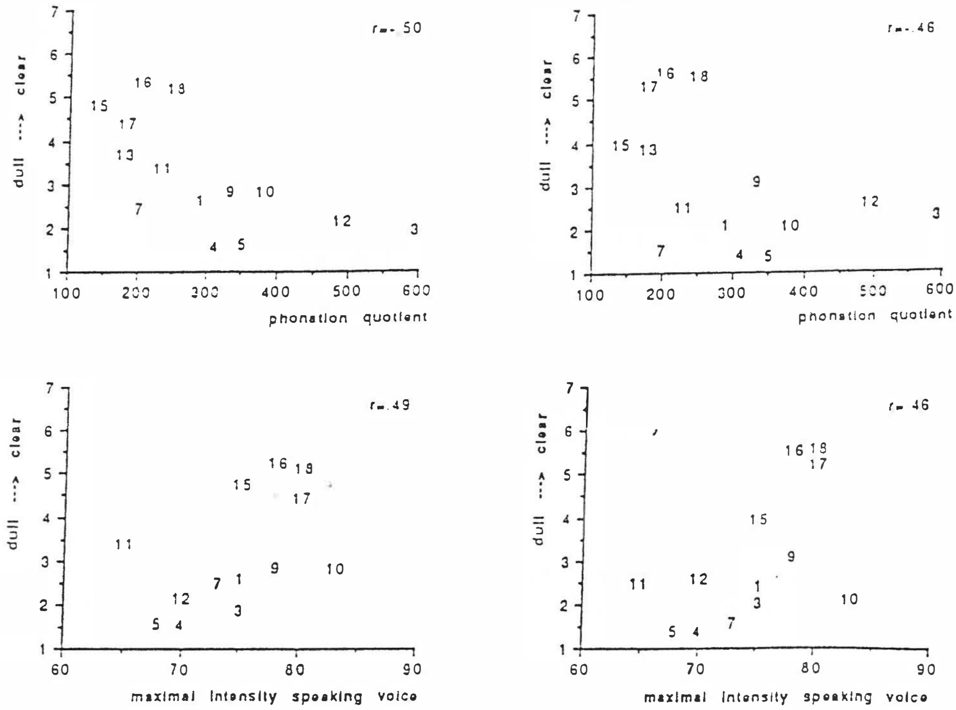

Figure 2 a. Scatterdiagrams of the scale *dull-clear* with the phonation quotient in ml/s (upper diagrams) and with the maximal intensity of the speaking voice in dB (lower diagrams). Means of judgements from students per speaker are given at the left, from the speech therapists per speaker at the right.
Patients before radiotherapy = 1,3,4,5 (no measurements for speaker 6)
Patients after radiotherapy = 7,9,10,11,12
Control speakers = 13,15,16,17,18 (max. intensity is not measured for speaker 13)
Corr. Coeff. are given in the upper righthand corner.

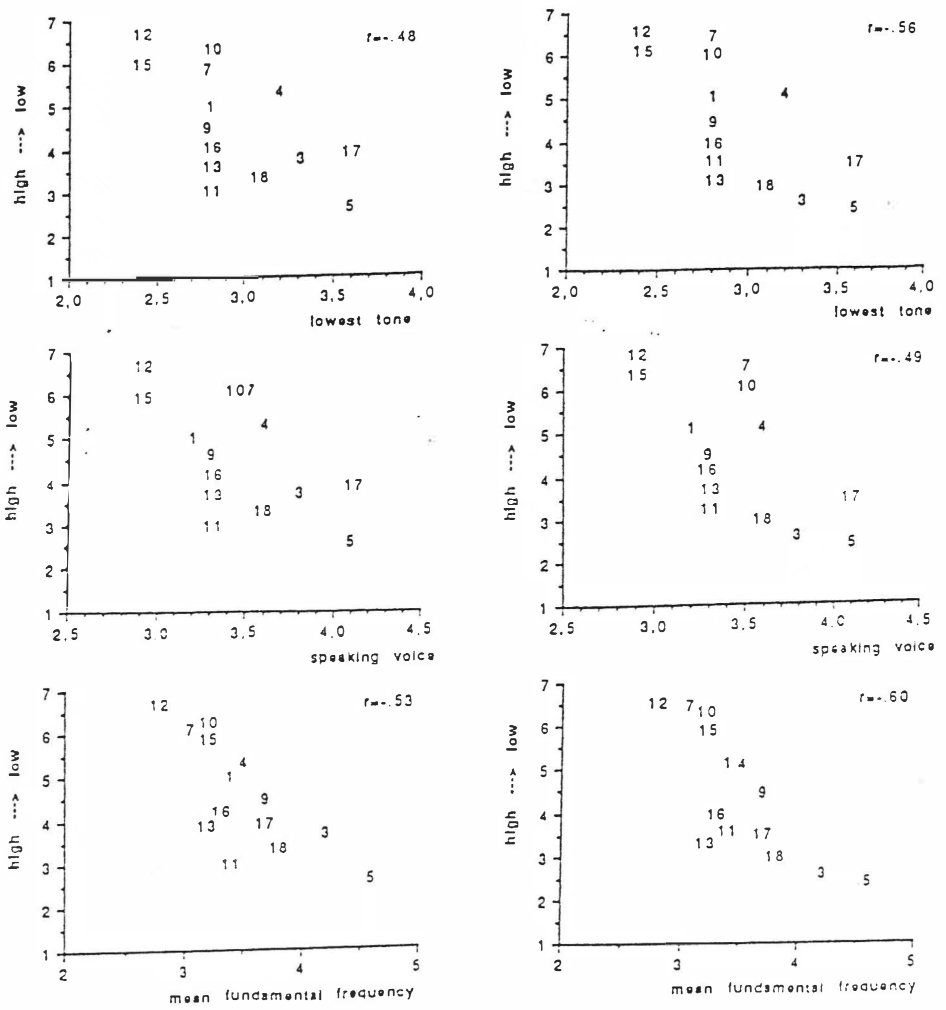

Figure 2 b. Scatterdiagrams of the scale *high-low* with the lowest tone (upper diagrams), with the speaking voice (middle diagrams), and with mean fundamental frequency (lower diagrams), all in semitones.
Means of judgements from students per speaker are given at the left, from the speech therapists per speaker at the right.
Patients before radiotherapy = 1,3,4,5 (no measurements for speaker 6)
Patients after radiotherapy = 7,9,10,11,12
Control speakers = 13,15,16,17,18
Corr. Coeff. are given in the upper righthand corner.

The correlations appear to be low. The highest correlation (r=.60) is the correlation between the scale *high-low* and the mean fundamental frequency, which is not quite surprising.
The diagrams show that the correlations are more or less linear. The diagrams for the students and the speech therapists are quite similar.

In the diagrams of the scale *dull-clear* the control group (speakers nrs 13,15,16,17,18) forms a separate group. The speakers in the control group are judged differently: e.g. speaker 15 has the same maximal intensity as speakers 1 and 3. Still, he is judged as *clear* and the others as *dull*. This pattern is found for speakers 16 and 9 as well.

The relation between the scale *dull-clear* and the phonation quotient is also different for the control speakers: speakers 16 and 7 have the same phonation quotient (both below the upper limit); still, speaker 7 is judged as *dull* and speaker 16 as *clear*.
These differences are found in the diagrams for the scale *high-low* as well, although here the control group is not a separate group. Speakers 10, 13, and 15 have the same mean fundamental frequency. Only speakers 10 and 15 are judged as *high*. The same counts for the scale *high-low* with the speaking voice (speakers 4 and 18) and with the lowest tone (speakers 10, 7 and 11, 13)

5. CONCLUSION

The results show that the general findings from literature are also found in this experiment. Pitch seems to correlate with mean fundamental frequency and with frequency data from the phonetogram (lowest tone and speaking voice).
Voice Quality is related to phonation quotient and maximal intensity.
There are also correlations between other dimensions and clinical parameters. Voice Appreciation correlates with phonation quotient and with maximal intensity of the speaking voice from the phonetogram. Articulation Appreciation is related to phonation quotient too. The dimension Tempo and the scale *intelligible-not intelligible* are not related to any of the clinical parameters.

However, the correlations appear to be low. The scatter diagams show that the correlations seem to be linear. They also show that the relations between the perceptual and clinical parameters are not simple. If two speakers have the same score on a clinical parameter, they can be judged differently by listeners. And, the other way around, if two speakers are judged the same, they do not necessarily have the same score in a clinical test.
The conclusion that relationships between the various parameters should be analysed by multivariate statistical techniques is also found in literature (Eskenazi et al., 1990; Moran & Gilbert, 1984; Kuwabara & Ohgushi, 1984; Murry et al., 1977; Hanson & Emanuel, 1979; Hammarberg et al., 1980). In further research, these techniques will be carried out.

Before these techniques will be carried out, there are also other aspects that will have to be examined. In this experiment the supposition was made that the semantic scales are interval scales. This supposition was based on previous research. Still, in this experiment the voices are pathological as well as normal, in contrast with the voices in the previous studies where then were all normal. If in further research, techniques will be used like factor analysis, regression analysis etc., first the nature of the scales used for judgements of pathological voices has to become clear.
Also, other scales can be added to the 15 scales which describe the abnormality of the pathological voices; but here too the nature of these scales has to be known. Further experiments can give a better insight to this dimension.

6. REFERENCES

Blom, J. & Herpt van L. (1976). "The evaluation of jury judgements on pronunciation quality", Proc. Inst. Phonetic Sciences, Univ. of Amsterdam 4, 31-46.

Buekers, P. (1980). "Veranderingen in de stemomvang gemeten met behulp van het fonetogram", Logop. Foniatr. 52, 132-145.

Eskenazi, L., Childers, D. & Hicks, D. (1990). "Acoustic correlates of vocal quality", Journal of Speech and Hearing Research 23, 298-306.
Fagel, W. & Herpt, L. van (1983). "Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation", Speech and Communication 2, 315-326.
Fritzell, B. & Hammarberg, B. (1977). "Clinical applications of acoustic voice analysis background and perceptual factors", IALPs 17th International Congress of Logopedics and Phoniatrics, 477-487.
Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg & J., Wedin, L. (1980). "Perceptual and acoustic correlates of abnormal voice qualities", Acta Otolaryngology 90,441-451.
Hanson, W. & Emanuel, F. (1979)."Spectral noise and vocal roughness relationschips in adults with laryngeal pathology", Journal of Communication Disorders 12, 113-124.
Herpt, van L. (1986). "Influence of rater's sex on voice and pronunciation assessment", Proc. Inst. Phonetic Sciences, Univ. of Amsterdam 10, 19-39.
Hirano, M. (1981). Clinical examination of voice. Disorders of human communication 5, New York.
Hirano, M. (1989). "Objective evaluation of the human voice: clinical aspects", Folia Phoniatr. 41, 89-144.
Kuwabara, H. & Ohgushi, K. (1984). "Experiments on voice qualities of vowels in males and females and correlation with acoustic features", Language and Speech 27 part 2, 135-145.
Moran, J. & Gilbert, H. (1984). "Relation between voice profile ratings and aerodynamic and acoustic parameters", Journal of Communication Disorders 17, 245-260.
Murry, T., Singh, S. & Sargent, M. (1977). "Multidimensional classification of abnormal voice qualities", J. Acoust. Soc. Am. 61, 1630-1635.