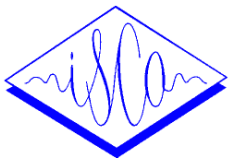

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

SPECTRAL CORRELATES OF BREATHINESS AND ROUGHNESS FOR DIFFERENT TYPES OF VOWEL FRAGMENTS

Guus de Krom

Research Institute for Language and Speech, University of Utrecht
Trans 10, 3512 JK Utrecht, the Netherlands

ABSTRACT

Breathiness and roughness ratings were related to a number of spectral parameters, including, among others, the relative peak level of the first harmonic, Harmonics-to-Noise Ratios (HNR) in selected frequency bands, and level differences between these frequency bands. Analyses were performed for 200 ms vowel onset fragments, 200 ms mid-vowel (post-onset) fragments, and 1000 ms fragments covering both the onset and post-onset parts of a vowel. HNR in the main energy frequency band was the best single predictor of both breathiness and roughness, explaining up to 55% of the variance. A combination of predictors explained 70% of the breathiness variance for all three types of fragments. For the roughness data, the same combination of predictors explained most of the variance in vowel onset fragments (61%), and least in post-onset fragments (35%). Thus, the onset seems to contain more acoustic information relevant to the perception of roughness than the mid-vowel fragment.

I. INTRODUCTION

In the literature on pathological voice quality research, several studies have been reported in which auditory impressions of voice quality, including breathiness and roughness, are related to acoustic or physiological parameters [1, 2, 3]. Yet, for a number of reasons, the question of which acoustic parameters may serve to describe the degree of breathiness and roughness severity and which of these parameters may be of use to discriminate between a breathy and a rough voice quality largely remains to be answered. Also, little is known about the possible influence of the type of voice fragment used for investigation.

In a previous experiment [4], it was found that roughness was rated more reliably for stimuli including the onset part of the vowel than for stimuli that consisted of the acoustically more stable mid-vowel segment only. These findings suggested that the onset of a vowel may contain additional perceptual cues with regard to the perception of certain voice quality aspects (at least roughness). Summarizing, the aims of this study were:

(1) to investigate which spectral parameters may serve as relevant predictors of breathiness and roughness, and

(2) to compare these findings for different types of vowel fragments.

II. METHODS

2.1 Subjects

Seventy-eight speakers were recorded, including 57 voice patients (women and men, suffering from different types and degrees of disorders). The 21 healthy speakers had no complaints about their voices. The listeners were six females, all third-year students of speech pathology.

2.2 Recording procedures

Recordings were made in a sound-isolated booth, using a condenser microphone. The speakers were asked to produce a number of sustained vowels /a:/ at conversational pitch and loudness. The vowels were band-pass filtered between 20 and 20,000 Hz, and stored on a DAT recorder (sf 48.0 kHz).

For each speaker, the experimenter selected one vowel that sounded most like the speaker's habitual, conversational voice. These vowels were low-pass filtered (9.6 kHz) and digitized at 12 bits (sf 20.0 kHz). Three different types of fragments were obtained from each recorded vowel; a vowel onset fragment, covering the initial 200 ms of the vowel, a 200 ms post-onset fragment, starting 500 ms after vowel onset, and a 1000 ms whole vowel fragment, starting at vowel onset. All 3 types of fragments were given linear ramped offsets of 12.5 ms. The post-onset fragments were given linear ramped onsets of 12.5 ms as well.

2.3 Perceptual evaluation

The 234 vowel fragments (78 speakers × 3 types) were presented over headphones in a sound-treated booth. The listeners were asked to evaluate all stimuli on a number of aspects (overall degree of deviance, breathiness, roughness, instability, voice weakness, and strain), using 10-point Equal-Appearing Interval scales, for which a rating of 1 was defined as not present, and a rating of 10 as maximally present. Breathiness was defined as a pathological, lax type of voice, associated with insufficient glottal closure, and roughness as a voice with a low-frequency noise component. Stimulus presentation was self-paced, and controlled by a computer program. Each fragment was

rated twice by each listener, in random order. The different types of stimuli were rated in separate listening sessions.

Next, the obtained voice quality ratings were analyzed by means of a multilevel analysis program [5], using a model for the analysis of variance with 3 random factors, namely the variance of the listeners' mean ratings, the variance of the speakers' mean ratings (i.e. the true score variance), and the replica variance. Rating reliability coefficients were determined on the basis of the relative magnitudes of the variance of the speakers' mean ratings and the variance of the means of the replicated ratings [4]. The reliability of roughness ratings was lower for the post-onset fragments (.79) than for the vowel onset and whole vowel fragments (.89 and .88). For breathiness, a less distinct fragment-type effect was found (.88 for post-onset, .90 for vowel onset, and .93 for whole vowels).

2.4 Spectral analyses

For each of the 234 vowel fragments, a number of spectral parameters were calculated, including the spectral level in four frequency bands: b0, 60 to 400 Hz; b1, 400 to 2000 Hz; b2, 2000 to 5000 Hz; b3, 5000 to 8000 Hz. Spectrum levels were defined as the base-10 logarithm of the summed power (squared magnitude) spectrum samples in a frequency band. Level differences between the frequency bands yielded spectral-slope parameters (LowSlope = $Level_{b0}$ - $Level_{b1}$; MidSlope = $Level_{b1}$ - $Level_{b2}$; HighSlope = $Level_{b2}$ - $Level_{b3}$). Spectral Harmonic-to-Noise Ratios in the four frequency bands ($HNR_{b0}$to HNR_{b3}) were calculated by means of a cepstrum-based technique [6]. An F_0 estimate was calculated in the cepstrum domain by locating the first rahmonic peak, resembling the (average) pitch period of the signal in the analysis window [7]. Two parameters representing the relative magnitude of the first harmonic were calculated: one by subtracting the peak level of the second from that of the first (h_1h_2), and another by calculating the difference between the peak level of the first harmonic and the level in the main energy band (h_1Level_{b1}). Analysis frames for which HNR_{b0} dropped below 5.0 dB were considered devoiced. In such cases, F_0, h_1h_2 and h_1Level_{b1} were given a missing value code. Finally, a parameter representing the percentage of devoiced analysis frames in a particular voice fragment (%devoiced) was determined.

Parameter values were calculated for each fragment by shifting a 1024-point Hanning window over 256 samples (12.8 ms), yielding 13 successive data points for each parameter for the 200 ms vowel onset and post-onset fragments, and 75 for the whole vowel fragments. The means and standard deviations of these 13 or 75 data points were treated as separate predictors in further analyses, and are identified by the prefixes m and s, respectively ($sHNR_{b0}$.therefore refers to the within-fragment standard deviation of HNR in the b0 band, rather than to the mean value, which is referred to as $mHNR_{b0}$).

2.5 Multilevel regression analyses

2.5.1 methods single predictor models.

Using the three-level models for the analysis of variance, the acoustic parameters were modelled as predictors of the true score variance. The percentage of variance explained (%EXP) was defined on the basis of the initial true score variance (INI, 100%), and the true score variance that remained after one of the acoustic parameters had been modelled as predictor (REM) (1):

$$\%EXP = (1 - (REM / INI)) \times 100\% \quad (1)$$

2.5.2 methods multiple predictor models

In order to determine which combination would yield the best results in the multiple predictor models, a factor analysis was performed on the correlation matrices of the acoustic parameters. The results indicated that the acoustic parameter spaces for each type of fragment could be described by six factors. The amount of variance accounted for by these six factors was 75.8% (vowel onset fragments), 75.2% (post-onset), and 78.5% (whole vowel). Based on their factor loadings and percentage of variance explained by the individual parameters, the following eight parameters were selected for entry in the analysis models: $mHNR_{b0}$, $mHNR_{b1}$, $mHNR_{b2}$, mh_1Level_{b1}, sLowSlope, mF_0, sF_0, and mHighSlope. The predictors were entered blockwise into the regression models. The output of these models consisted of the remaining variance estimates, an intercept, and regression coefficients for the predictors. A two-tailed 5% significance level was adopted for the estimated regression coefficients. Predictors whose regression coefficients did not meet this criterion were dropped, after which new iterations were run. This purging process was repeated until all regression coefficients met the significance criterion. The percentage of true score variance explained was determined as in (1).

III. RESULTS

3.1 Single predictor models

For each one of the predictor variables, the percentage of true rating variance explained was calculated as in (1). Results for breathiness and roughness are given in Table 1.

Table 1. Percentage of true rating variance explained by the acoustic parameters. Data are given only for parameters that explain at least 20% of the variance. The signs indicate whether the correlation between the acoustic parameter and the voice quality aspect is positive or negative. Results are given for vowel onset (VO), post-onset (PO), and whole vowel fragments (WV). Breathiness data are given in the left hand columns (B); roughness data are given in brackets in the right hand columns (R).

	VO		PO		WV	
	B	(R)	B	(R)	B	(R)
mLevel$_{b1}$	-20		-30			
mLowSlope	+21	(+20)	+41	(+27)	+25	(+24)
mMidSlope		(-34)				
mHNR$_{b0}$	-26	(-42)		(-21)	-21	(-24)
mHNR$_{b1}$	-44	(-55)	-44	(-32)	-48	(-35)
mHNR$_{b2}$	-37	(-25)	-39	(-25)	-42	(-23)
mh$_1$h$_2$			+21			
mh$_1$Level$_{b1}$	+24		+37		+26	
sLevel$_{b1}$			+21			
sHNR$_{b0}$					+24	(+26)
sHNR$_{b2}$	-23		-27		-21	
sLowSlope	+31		+24			
%devoiced	+29	(+29)			+45	(+27)

As can be observed, few parameters explained more than 40% of the rating variance. Mean HNR in the lower two frequency bands (b1 and b2) proved among the best predictors of breathiness and roughness for all three types of fragments, mHNR$_{b1}$ explaining 55% of the roughness variance in vowel onset fragments. The parameters reflecting the level of the first harmonic (mh$_1$h$_2$ and mh$_1$Level$_{b1}$) were useful predictors of breathiness, but not of roughness. The percentage of devoiced frames in the fragment (%devoiced) proved a useful predictor of breathiness and roughness in onset and whole vowel fragments.

Most s parameters explained less than 20% of the variance. sHNR$_{b2}$ was the only parameter to explain more than 20% of the breathiness variance in all three types of fragments. sHNR$_{b0}$ explained just over 20% in whole vowel fragments. sLowSlope explained up to some 30% of the breathiness rating variance in vowel onset fragments, and just over 20% in post-onset fragments.

3.2 Multiple predictor models.

The results for the multiple predictor models are given in Table 2.

Table 2. Standardized regression coefficients for acoustic parameters in the final analysis models. %EXP = percentage of true variance explained. Blanks were used for coefficients that did not fulfil the 5% significance criterion. Results are given for vowel onset (VO), post-onset (PO), and whole vowel fragments (WV). Breathiness data are given in the left hand columns (B); roughness data are given in brackets in the right hand columns (R).

	VO		PO		WV	
	B	(R)	B	(R)	B	(R)
mF$_0$	.76	(.42)	.76		.91	
sF$_0$		(.28)	.20			(.61)
sLowSlope	.46		.37		.36	
mHNR$_{b0}$	-.37	(-.53)	-.35		-1.05	(-.69)
mHNR$_{b1}$	-.74	(-.98)	-.66	(-.64)		
mHNR$_{b2}$			-.30	(-.31)	-.44	
mh$_1$Level$_{b1}$	.28		.51		.50	(.62)
mHighSlope	-.50	(-.26)	-.36		-.38	
%EXP	68	(61)	69	(35)	68	(43)

As can be observed, the three mHNR parameters correlated negatively with rated breathiness and roughness, indicating that a decrease of harmonic energy in frequency bands up to 5 kHz was associated with a breathy or rough voice quality. A relatively high level of the first harmonic and a relatively high level of frequency components above 5 kHz also contribute to perceived breathiness, as indicated by the signs of the regression coefficients for mh$_1$Level$_{b1}$ and mHighSlope. For roughness, these parameters were less important predictors. As expected, the regression coefficients for mF$_0$, sF$_0$, and sLowSlope were all positive.

The percentage of variance explained is about equally high for all three breathiness models (almost 70%), although the model for post-onset fragments includes all eight predictors, compared to six for the vowel onset and whole vowel fragments. For roughness, the percentage of roughness rating variance explained is much higher for vowel onset fragments (61%) than for whole vowel fragments (43%) and especially post-onset fragments (35%). Consequently, different predictors appear in the three models, although each model contains at least one spectral noise related parameter.

IV. DISCUSSION AND CONCLUSIONS

The results of the single predictor analyses indicated that none of the acoustic parameters could be considered an outstanding predictor of either rated breathiness or roughness in this study. mHNR$_{b1}$ and mHNR$_{b2}$ ranked among the better predictors of both breathiness and roughness severity for all three types of fragments.

Based on previous studies [1, 2, 3], it was expected that the high-frequency spectral slope, the relative level of the first harmonic, and the (mean) Harmonics-to-Noise Ratio in higher frequency bands would prove viable predictors of rated breathiness. However, the high-frequency slope of the spectrum explained little variance. On the other hand, the data confirmed that breathiness is associated with a relatively high first harmonic. Roughness rating variance could best be related to measures of spectral noise and the percentage of devoiced frames in the signal fragment. As expected, parameters related to the relative level of the first harmonic proved less useful predictors of roughness than of breathiness. Fundamental frequency, and, to a lesser extent, the overall intensity of the signal, proved poor to very poor predictors of rated breathiness or roughness, which suggests that our listeners had not followed a naive listening strategy, but that they had based their ratings of breathiness and roughness severity on other, more subtle acoustic cues instead.

The amount of breathiness rating variance explained by the multiple predictor models was about 70% for all three types of fragments, which is substantially higher than the 48% of variance explained by the best single predictor model, indicating that the perception of breathiness can be related to several spectral characteristics, rather than to one single spectral feature. A lowered Harmonics-to-Noise Ratio was an important predictor of both breathiness and roughness. The data for roughness in vowel onset fragments (which yielded by far the best model of all three types of fragments) indicated that roughness was associated with low HNR values in frequency bands up to 2 kHz. The emergence of spectral noise in the 2 to 5 kHz band was more typical of breathiness. Thus, some evidence was found that the frequency distribution of spectral noise components may be of help to distinguish between breathiness and roughness. The high-frequency spectral slope and the relative peak level of the first harmonic proved more viable predictors of breathiness than of roughness. The s parameters that reflected the frame-to-frame fluctuation of parameter values generally showed a higher correlation with breathiness than with roughness. This result was considered a bit surprising, because an irregular or unstable nature of the signal is more usually associated with a rough than with a breathy voice quality.

Despite the fact that the regression models for the breathiness and roughness data exhibited typical differences, the data do not indicate that the spectral parameters that were examined allow for a clear-cut distinction between breathiness and roughness. Part of this may be explained on the basis that the speakers recorded for this study often exhibited both breathy and rough aspects in their voices [4]. Besides, it may be that breathy and rough voices truly do not differ that much in terms of acoustic properties. Breathiness and roughness are, after all, highly related phenomena in a number of ways. The acoustic differences between breathy and rough voices may, in other words, actually be as subtle as they appear to be in this study.

Whereas the three breathiness models each explained 68% of the rating variance, the percentage of roughness rating variance that could be explained on the basis of the selected spectral parameters was generally much lower. The vowel onset model explained most (61%), followed by the whole vowel model (43%), and the post-onset model (35%). It may be interesting to compare these results to the roughness rating reliability coefficients. The relatively low percentage of variance explained by the post-onset model agrees with the relatively low roughness rating reliability for post-onset fragments (.79). However, the difference in the percentage of variance explained by the vowel onset and whole vowel models is not reflected in the rating reliability data, as both types of fragments were rated about equally reliably (.89 [vowel onsets]; .88 [whole vowels]). We therefore assume that the better fit of the vowel onset model as compared to the whole vowel model has its basis in acoustic differences between the two types of fragments. Apparently, the vowel onset contains more information that may be relevant for the perception of roughness than the acoustically more stable mid-vowel segment. In addition, differences between breathy and rough voices may relate to the timing of acoustic events, rather than to the nature of these events themselves. Acoustic disturbances that primarily occur during the onset of voicing would then be associated with roughness, whereas phenomena that last throughout a vowel would give rise to a breathy sensation..

REFERENCES

[1] Childers, D.G., & Lee, C.K. (1991). Voice quality factors: Analysis, synthesis, and perception. JASA, 90, 2394-2410.

[2] Hammarberg, B. (1986). Perceptual and acoustic analysis of dysphonia. Stockholm: Dissertation Department of Logopedics and Phoniatrics, Huddinge University Hospital.

[3] Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. JASA, 87, 820-857.

[4] De Krom, G. (in press). Consistency and reliability of voice quality ratings for different types of speech fragments. JSHR.

[5] Prosser, R., Rasbash, J., & Goldstein, H. (1991). ML3-software for three-level analysis. Users' guide for V.2. London: University of London, Institute of Education.

[6] De Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. JSHR, 36, 254-266.

[7] Noll, A.W. (1967). Cepstrum pitch determination. JASA, 41, 293-309.