IRREGULARITIES IN THE VOICE: SOME PERCEPTUAL EXPERIMENTS USING SYNTHETIC VOICES.

Jan Gauffin and Svante Granqvist, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Stockholm, Sweden
Britta Hammarberg, Stellan Hertegård and Alf Håkansson, Dept. of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital, Huddinge, Sweden

ABSTRACT

Sustained vowels with different kinds and magnitudes of jitter and shimmer sequences were synthesised and evaluated in a listening test. Though having the same jitter and shimmer magnitude, the different sequences were rated differently.

INTRODUCTION

For normal voices the acoustic effect of irregularities in the voice source is, in most cases, subtle, but pathological changes in the larynx may cause it to be a prominent feature. As we can hear rather small perturbations in the voice, we should also be able to measure and make objective classification of such perceptual voice qualities. Unfortunately, most attempts to do so have been rather disappointing.

Methods of chaos physics can be used to explain the mechanism behind observed irregularities in the voice source such as period doubling and chaotic vibrations of the vocal folds [1]. Simple simulation models of the vocal folds, like the two-mass model, can also be used to illustrate these mechanisms. However, understanding the mechanisms behind the irregularities in the voice source does not automatically give us the methods to analyse the corresponding acoustic features.

One kind of irregularity in the voice is the occurrence of more or less regular patterns of period-to-period variability. Such voices are often referred to as rough or creaky [2]. In the frequency domain, this corresponds to subharmonics in the spectrum. In the present investigation we have been studying the perception of synthetic voices with subharmonics by using synthetic vowels with repetitive patterns of perturbed periods.

METHOD

The synthesis was produced using the LF-model by Fant, Liljencrants & Lin [3], followed by a vocal tract filter tuned for the vowel /a/ [4]. Average fundamental frequency was 100 Hz and sampling frequency 16000 Hz. Different sequences of fundamental periods were generated, following the sequences AB, AOBO and AABB (see Tables 1 and 2). In the jitter case, A stands for a prolonged period and B for a shortened one, whereas O stands for an unmodified one. In the shimmer case, A stands for a period with higher amplitude and B stands for a period with lower amplitude.

This results in six different types of stimuli. Each of these was generated with 10 different magnitudes giving a total of 60 stimuli.

The sequences AOBO and AABB should give the same average jitter or shimmer measure since the AOBO sequence varies 1 unit between every period and the AABB sequence varies 2 units every other period. The AB sequence varies 2 units between every period. Therefore the magnitude of the peak period-to-period variability was divided by 2 in the AB sequence. The different sequences are illustrated in Figure 1 and the spectra of the six types of stimuli in Figure 2. All jitter and shimmer values refer to the voice source.

Table 1. The jitter sequences, period time deviations, n=0,1,..,9

Per. #	AB	AOBO	AABB
1	+n·0.6%	+n·1.2%	+n·1.2%
2	-n·0.6%	0%	+n·1.2%
3	+n·0.6%	-n·1.2%	-n·1.2%
4	-n·0.6%	0%	-n·1.2%

Table 2. The shimmer sequences, period amplitude deviations, n=0,1,..,9

Per. #	AB	AOBO	AABB
1	+n·1%	+n·2%	+n·2%
2	-n·1%	0%	+n·2%
3	+n·1%	-n·2%	-n·2%
4	-n·1%	0%	-n·2%

The equal average period-to-period variability should make it possible to examine and compare the perceptual sensitivity between the different sequences. However, the jitter and shimmer sequences with the same n are not comparable, since the step magnitudes of 0.6% and 1% are chosen arbitrarily.

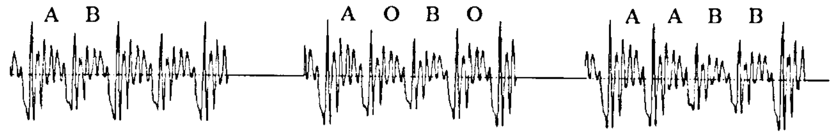

Figure 1. The three different shimmer sequences. n=9 (Table 2).

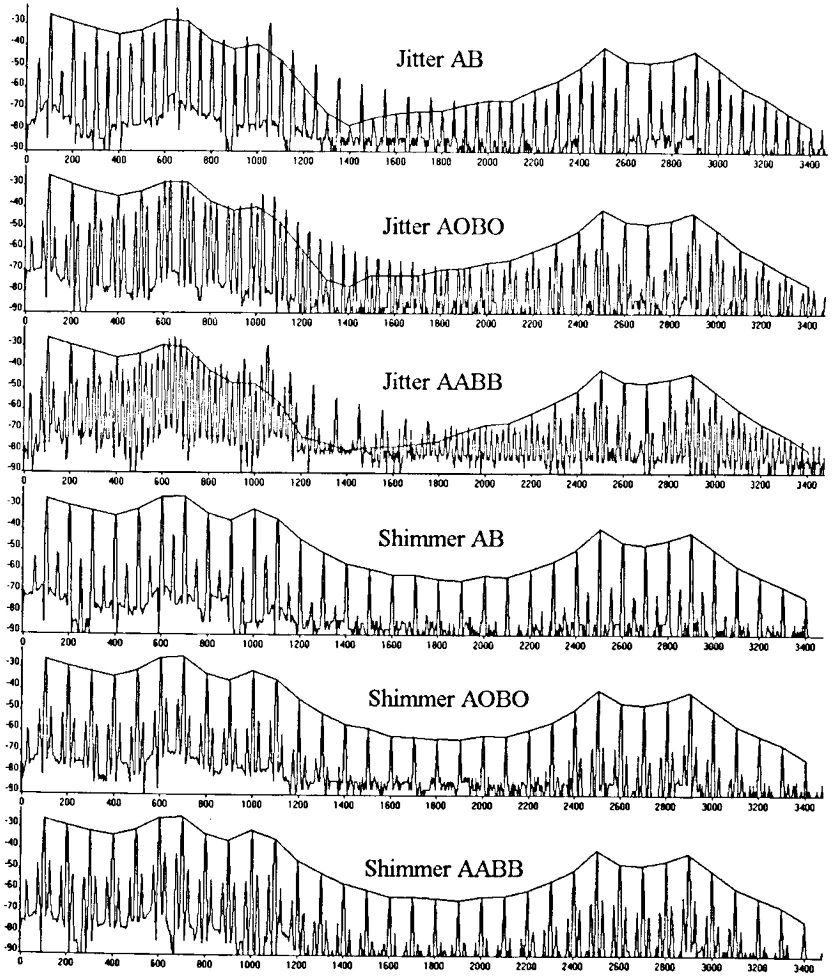

Figure 2. Narrow-band spectra of the six different types of stimuli and the envelope of harmonics. n=6 (Tables 1&2).

In Figure 2, one may note the difference between the jitter and shimmer stimuli by the envelope of the subharmonics. The jitter stimuli contain much weaker low-frequency subharmonics than higher-frequency subharmonics whereas the shimmer subharmonic envelope follows the envelope of the harmonics. It is also obvious that the stimuli with AABB or AOBO sequences have 25 Hz between spectral peaks, while the AB sequences have 50 Hz between the peaks. (Strictly speaking, the fundamental frequency has dropped from 100 Hz to 50 or 25 Hz, but if the subharmonics are weak enough the perceived pitch will still be 100 Hz). The spectra of AOBO and AABB stimuli appear similar but not identical.

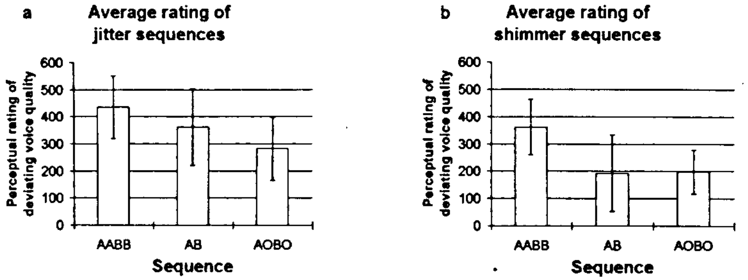

Figure 3a, b. Average and standard deviation of ratings of jitter (a) and shimmer (b) sequences. Each bar is an average of ratings of n=0,1,..,9 and all listeners.

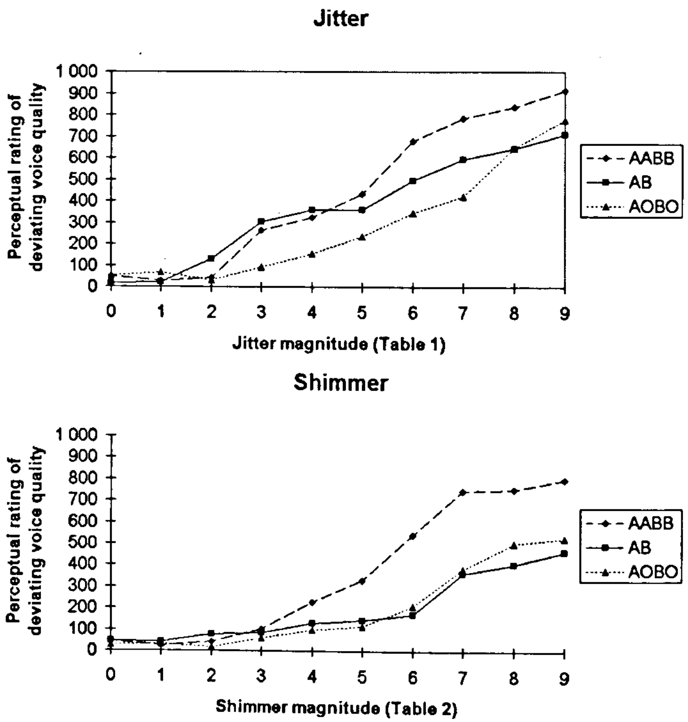

Figure 4. Average of all listeners' ratings of the different sequences.

The 60 stimuli were presented in random order to two experienced voice clinicians, two voice researchers and one musician. Each listener listened to the stimuli between one and three times, giving a total of 12 listening series. The listeners were asked to rate the degree of deviating voice quality on a visual analogue scale [5]. The position on the analogue scale was translated to a number between 0 (no deviation) and 1000 (maximum deviation).

RESULTS AND DISCUSSION

The results indicated that the sequences were rated quite differently regarding deviating voice quality. In the jitter case, the AABB sequence was rated 20% higher than the AB sequence and the AOBO was rated 22% lower than the AB sequence on average (see Figure 3a). In the shimmer case, the AABB sequence was rated 88% higher than the AB sequence and the AOBO was rated 3% higher than the AB sequence on average (see Figure 3b).

Figure 4 shows the perceptual ratings as a function of the magnitudes of jitter and shimmer for the different sequences. As expected, an increased degree of jitter or shimmer results in an increased perceptual rating of deviating voice quality.

A closer look at the spectra for the (Figure 2) might explain why the different types of stimuli were rated differently. Comparing the AABB and AOBO spectra, the latter have slightly weaker subharmonics and, furthermore, the jitter AOBO spectrum lacks the subharmonics 0.5·F0, 1.5·F0 etc.

This indicates that the levels of subharmonics could be more perceptually relevant than the jitter and shimmer measures.

All the listeners rated the AABB and AOBO stimuli as above. However they disagreed on how the AB sequence should be rated compared to AABB and AOBO sequences. This can be seen in the standard deviation bars in Figure 3. The AABB and AOBO sequences had a similar sound quality, while the AB stimuli differed from them. The different sound qualities can be explained by differences in the spectra. The AB sequence had 50 Hz between spectral peaks and the AABB and AOBO had 25 Hz. This might explain why the listeners disagreed on how to compare the AB stimulus with the AABB or AOBO stimuli.

CONCLUSIONS

We have shown that different kinds of jitter and shimmer sequences with the same average period-to-period variability are rated differently. This indicates that methods using period-to-period variability as a way to rate voice qualities might fail. We have also seen that there are differences in narrow-band spectra of the stimuli that might explain the differences in perceptual rating of the stimuli. For voices with repetitive patterns in period-to-period variability, this suggests that a method analysing spectral characteristics might yield better results than jitter or shimmer methods.

ACKNOWLEDGEMENT

This work was supported by research grants from the Bank of Sweden Tercentenary Foundation and the Karolinska institute.

REFERENCES

[1] Lauterborn, W. & Parlitz, U. (1988): Methods of chaos physics and their application to acoustics. *J. Acoust. Soc. Am.*, vol. 84, pp. 1975-1993.

[2] Imaizumi, S. (1986): Acoustic measures of roughness in pathological voice. *J. Phonetics*, vol. 14, pp. 457-462.

[3] Fant, G., Liljencrants, J. & Lin, Q. (1985): A four-parameter model of the glottal flow. *STL-QPSR*, vol. 4, pp. 1-13 (Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm).

[4] Håkansson A. (1995) *LF-Edit. Windowsprogram för LF-modellen.* Manual for custom-made program. Dept of Logopedics and Phoniatrics, Karolinska institute, Huddinge University Hospital

[5] Wewers, M.E., Lowe, N.K. (1990): A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, vol. 13, pp. 227-236.