Jitter in the Singing Voice

R.A. Rasch
Utrecht, the Netherlands

1. Introduction

In the musical tones produced by the singing voice, frequency variations are always present. They can be grouped into three categories: (1) *trend* (slow, gradual changes), (2) *vibrato* (periodical variations with a frequency of 5 to 6 Hz mostly), and (3) *jitter*, to be defined as small-scale period-to-period fluctuations with a random or pseudo-random character. The vibrato of the singing voice has been studied extensively, beginning with Seashore (1932), but this is not the case with jitter. Jitter has been studied in connection with pathological speech (Lieberman, 1963) and from a point of view of hearing theory (i.e. Pollack, 1968; Cardozo and Ritsma, 1968), but only very few studies are available that deal with the jitter of any musical tones, and they are mostly concerned with the tones of stringed instruments (e.g., Cremer, 1973 and McIntyre et al., 1981). A few data about the singing voice are given by Bennett (1981). As far as my knowledge goes, the jitter of the singing voice has not been studied systematically. Still, jitter is present in the singing voice to a significant extent, as will be shown in this paper.

Jitter (s_j) is quantitatively defined as the standard deviation of period duration (s_p) when there is no vibrato or trend:

$$s_j = s_p = (\Sigma (p_i-\bar{p})^2 / n)^{1/2},$$

in which p_i is the duration of the i-th period, p the mean period duration, and n the number of periods. This measure can be made more meaningful in a musical context by interpreting the standard deviation as a musical interval which can be converted into cents*:

* The *cent* scale is a logarithmic transformation of the interval as frequency ratio:

$$I = 1200 \log_2 q/p \text{ cents},$$

where q/p is the frequency ratio, and I the interval size in cents. One cent is 1.000578/1, the octave (2/1) is 1200 cents, a semitone in equal temperament ($2^{1/12}/1 = 1.059$) is 100 cent.

$$J = 1200 \log_2 (1+s_j/\bar{p}) \text{ cents},$$

which measure we will use throughout this paper.

When vibrato and trends are also present, the situation is a bit more complex. The amount of jitter (s_j, now defined as the standard deviation of period durations as far as due to jitter) can be derived from the variances of period durations (s_p^2) and of the time differences of successive periods (s_d^2) in the following way:

$$s_j^2 = \frac{(1-r_v)s_p^2 - \frac{1}{2}s_d^2}{r_j - r_v},$$

in which r_j and r_v are the autocorrelations of successive period durations due to jitter (r_j) and to vibrato/trend (r_v), respectively. For a vibrato, this autocorrelation equals cos $(2\pi f_v/\bar{f})$, in which F_v is vibrato frequency and $\bar{f}$ is mean signal frequency. When vibrato/trend is sufficiently slow compared to the period durations and the variance due to vibrato/trend is not too large, then the amount of jitter is well approximated by:

$$s_j = \frac{s_d}{(2 - 2r_j)^{1/2}},$$

which means that, when r_j can be assumed to be constant, jitter is about proportional to the standard deviation of the time differences of successive periods. For the singing voice, r_j can be assumed to be about zero.

2. Measurement Procedure

Musical tones sung on Dutch syllables in an anechoic room were recorded on tape. Period durations were measured be feeding the filtered fundamental component to a Schmitt trigger that produced pulses at every positive zero crossing. The onset times of the pulses were measured by a 10 MHz clock and stored in computer memory. Further analyses were done with help of computer programs. Diagrams with instantaneous frequency (as the reciprocal of period duration) could be plotted on a graphic terminal (see Fig. 1).

3. Results

We will present here the results of measurements of the jitter in *h*-vowel-*t* syllables sung by two professional singers, a bariton and a soprano. They sang nine different vowels on various fundamental frequencies in three modes: straight (no vibrato), neutral (small vibrato), and strong vibrato. Mean jitter of the nine vowels (i.e. averaged per vowel over the three singing

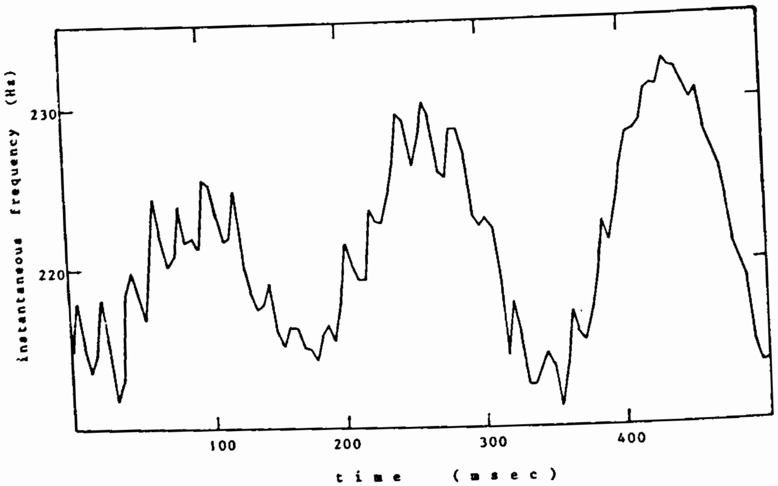

Figure 1. Frequency per period as a function of time. The tone is A2 (220 Hz), sung with vibrato by the bariton. The duration of the fragment is 500 msec. The number of periods is 110. The vowel is /u/ in Dutch *hoet*, with an amount of jitter of about 10 cents and a vibrato depth of about 60 cents.

modes) has been presented as a function of vowel in Figure 2, with signal frequency as parameter. Generally spoken, there is a frequency effect. Low tones have more jitter than high tones. But there is a vowel effect as well. Vowels with a low first formant frequency F_1 (like Dutch *hiet* and *huut*) show relatively small amounts of jitter, while vowels with a high F_1 (like Dutch *haat* and *hat*) show relatively large amounts of jitter.

I want to propose here the hypothesis that the jitter differences between vowels are due to the interaction between the vocal cords and the vocal tract. The vocal cords and the vocal tract may be seen as two coupled oscillators. In principle, the vocal cords are driving, while the vocal tract is driven, which implies a certain phase relation between the oscillations of the two systems. However, because of the coupling the vocal tract may also have a driving impact on the vocal cords, which implies a phase relation that is surely not the reverse of the one just mentioned. It seems plausible that these conflicting phase relations can be the cause of a small phase instability in the transmitted wave from the vocal cords to the vocal tract, which, as a matter of fact, becomes apparent in the produced sound signal as a small frequency instability.

The phase lag ϑ of a driven oscillator relative to a driving one is most simply given by $\cot \vartheta = Q(F_1/F_0\text{-}F_0/F_1)$, in which Q refers to the damping of the driven oscillator, F_0 is the frequency of the driving oscillator and F_1 is the resonance frequency of the driven oscillator. When we use the F_1 values given by Pols et al. (1973) and Nierop et al. (1973), the correlation between the amount of jitter and the quantity $F_1/F_0\text{-}F_0/F_1$ appeared to be 0.89 for the 131 and 220 Hz bariton tones and 0.85 for the 220 Hz soprano tones. For the

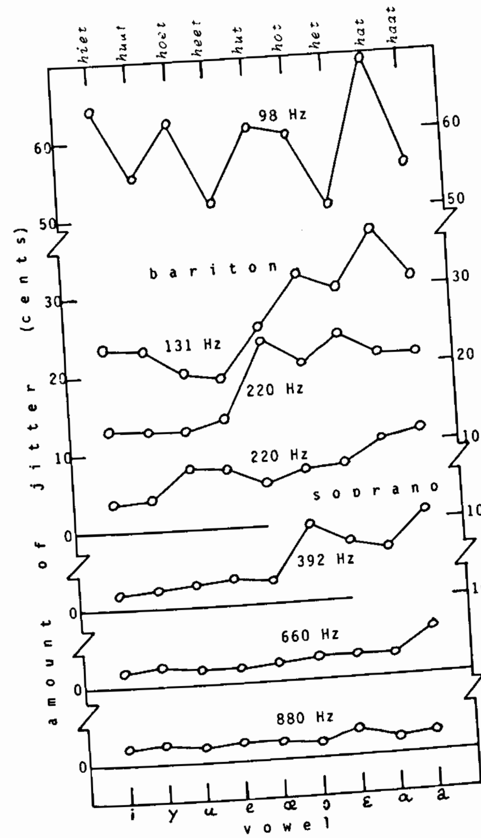

Figure 2. Amount of jitter as a function of vowel. Parameter is signal frequency. The vowels are ordered after their first formant frequencies, from low to high. We used data given by Pols & al. (1973) and Nierop & al. (1973).

highest soprano tones, the calculation of such a correlation makes less sense, since the F_0 then is in the neighbourhood of the F_1, while it is possible that the singer adapted the F_1 in one way or another to the F_0. The data show that, as long as F_0 is around or above F_1 the amount of jitter is at some minimum base-line level of about 2 to 3 cents*. The vowels with an F_1 substantially (more than 100 to 200 Hz) higher than the F_0 have more jitter. For the 392 Hz tones, this is beginning with *hot* (F_1 578 Hz), for the 660 Hz tones only with *haat* (F_1 986 Hz), for the 880 Hz tones not at all.

The large, vowel-independent amounts of jitter of the 98 Hz bariton tones may be due to the surpassing of some critical frequency separation between F_0 and F_1.

4. Conclusion

The data given about the amounts of jitter present in the singing voice must be seen as preliminary. However, they are encouraging for further research. Jitter is a 'noisy' phenomenon, which can vary with singer, mode of singing, frequency and vowel, but also with attention and concentration, fitness and fatigue, exercise and mood. Really smooth, nice data should never be expected. However, jitter seems to be a rather fundamental property of musical tones in general. Couplings like the one between the vocal cords and the vocal tract are present in almost all acoustical musical instruments, like the strings and the top plate or sound board of stringed instruments, the reed and tube in the reed wind instruments, the lips and tube in brass wind instrument, etc. Very often one of the oscillators is vibrating with another than its resonance frequency. When jitter is indeed caused by the interaction between these coupled systems, it should be expected to be present in almost every musical tone. It may well be one of the features of tones that make possible a distinction between the dead sounds of electronic frequency generators and the live sounds of musical instruments.

Acknowledgements

My thanks go to Gerrit Bloothooft who lent me his recordings of singing-voice tones, and to the Audiology Department of the Institute for Perception TNO, Soesterberg, the Netherlands, where I could make use of computing and audio equipment.

References

Bennett, G. (1981). Singing synthesis in electronic music. In: *Research aspects on singing*. Royal Swedish Academy of Music, Stockholm. 34-50.

Cardozo, B.L., and Ritsma, R.J. (1968). On the perception of imperfect periodicity. *IEEE Trans. Audio Electroacoust.* **16**, 159-164.

Cremer, L. (1974). Der Einfluss des "Bogendrucks" auf die selbsterregten Schwingungen der gestrichenen Saite. *Acustica* **30**, 119-136.

Lieberman, Ph. (1963). Some acoustic measures of the fundamental periodicity of normal and pathological larynges. *J. Acoust. Soc. Amer.* **35**, 344-353.

McIntyre, M.E., Schumacher, R.T., and Woodhouse, J. (1981). Aperiodicity in bowed-string motion. *Acustica* **49**, 13-32. Corrigendum in *Acustica* **50** (1982), 294-295.

Nierop, D.J.P. van, Pols, L.C.W., and Plomp, R. (1973). Frequency analysis of Dutch vowels from 25 female speakers. *Acustica* **29**, 110-118.

Pollack, I. (1968). Detection and relative discrimination of auditory jitter. *J. Acoust. Soc. Amer.* **43**, 308-315.

Pols, L.C.W., Tromp, H.R.C. and Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *J. Acoust. Soc. Amer.* **53**, 1093-1101.

Rasch, R. (1983). Jitter in violin tones. Submitted to the Proceedings of the Stockholm Music Acoustics Conference 1983.

Seashore, C.E. (ed.) (1932). The vibrato. *Studies Psychol. Music* **1**. Iowa Univ., Iowa City.