# Qualities of a voice emeritus

Gerrit Bloothooft and Peter Pabon

Utrecht University


## Abstract

The effects of vocal ageing are investigated in a professional mezzo-soprano singer, for which the phonetogram, and 45 vowels, each sung at fundamental frequencies of 220, 392, and 659 Hz, were recorded at the age of 52 and 74 years. The comparison demonstrates a serious loss in the vocal range, dynamics and control: (1) a loss of half an octave in the highest fundamental frequency range, (2) a loss of 6 dB at the highest vocal intensities, (3) less accuracy in targeting of $F_0$, (4) no significant change in average vibrato frequency, but (5) much more instability in vibrato frequency and less vibrato modulation depth. The analysis of vocal vibrato is realized with a new method that allows computation of instantaneous vibrato frequency and extent (modulation depth).


## 1 Introduction

When a person comes of age, the vocal apparatus will not escape from physiological changes such as dehydration of tissues, ossification of cartilages, changes in muscular structure, reduction of the number of nerve fibers and reduction of the speed of action potentials, among others. A major acoustic effect of this ageing process is the gradual increase of the pitch of the speaking voice for males from about 120 Hz at the age of sixty, to 155 Hz (vowel /a/) or even 187 Hz (vowel /i/) for elderly over ninety years of age. Not so much change was found for females (Decoster, 1998), implying that male and female speaking voices become to some extent more similar at older age. In addition, Decoster reported an increasing acoustic instability, exemplified by an increase of the variability of the speaking fundamental frequency ($F_0$), an increase of jitter and shimmer (random fluctuations in frequency and amplitude of $F_0$), and a decrease of the harmonics-to-noise ratio of the spectrum. Spectral (envelope) changes over age seem to be less pronounced.

The lowest and highest possible fundamental frequency a voice can produce critically depends on the state of the vocal apparatus, specifically on subglottal pressure and muscular tension. As a consequence, it may be expected that the vocal tonal range is sensitive to ageing. And indeed, a decrease of the total fundamental frequency range sets in at the age of sixty, for both males and females, and reduces from about 24 semitones (two octaves) to 18 semitones at the age of ninety (Böhme & Hecker, 1970).

Whereas relatively little research has been devoted to the effects of ageing on the normal human voice, even less attention has been given to the special case of the professional singing voice. Singers optimally train their voice during their career, but of course at some moment also for them age will take its toll. For singers, the ageing process may be even much more critical than for non-singers, since they usually explore their voice to its dynamic and tonal extremes. Also voice control, the ability to follow precisely the musical score, is of eminent importance to a singer, and its precision may reduce with age. These capabilities are less important in speech, where less vocal control or reduced pitch range will longer go unnoticed, but for singers the slightest loss may foreshadow the end of a career.

In this study we explore the effects of age on the acoustic properties of a professional mezzo-soprano singer's voice. In 1981, recordings were made in the framework of a doctoral thesis on the spectrum and timbre of sung vowels (Bloothooft, 1985). At that time the singer was 52 years of age. We were in the circumstances that the recordings could be repeated 22 years later, in 2003, when she was 74 years old. This opened the possibility of a longitudinal study of acoustic changes in the same voice. We concentrated on influences on vocal dynamics, as measured in the phonetogram, and on vocal precision and control, by means of an analysis of vibrato in sung vowels. For the latter study a new method was developed for the decomposition of the pitch trace into transient effects, drift, vibrato, and jitter.

## 2 Recordings

In 1981, recordings were made of professional singers, among which the mezzo soprano who is central in this study. In an anechoic room at the Free University of Amsterdam, the nine vowels /a, ɑ, i, u, ɔ, œ, y, ɛ, e/ were sung in /h/-vowel-/t/ context with a duration of one to two seconds each, at fundamental frequencies ($F_0$) of 220 Hz (A3 or a), 392 Hz (G4 or g') and 659 Hz (E5 or e"). The singers were asked to sing these vowels in several modes, out of which the following five were used in the present study: neutral, light, dark, soft and loud. Recordings were made with a microphone positioned at 0.3 m from the singer. The recordings in 2003 were made in the personal studio of the singer. A microphone pair, one at a distance of 0.3 m, the other close to the mouth, was fixed on a headset. The microphone close to the mouth was used for the softer phonations (but calibrated for the sound level at 0.3 m), while the other microphone was used for the loud phonations. In this way an optimal signal-to-noise ratio could be achieved.

The phonetogram recordings from 1981 were the first using a computerized measurement technique (Bloothooft, 1981). Under the same conditions as described above, the singer stood facing a monitor on which $F_0$ was presented on the x-axis and vocal intensity (I) on the y-axis. Real-time measurement of $F_0$ and I was translated to a position of a marker on the monitor. With the help of this visual feed-back the singer was asked to complete the phonetogram, i.e. to sing all possible combinations of $F_0$ and I. In subsequent separate sessions, the singer was asked to make a registration of a phonetogram while singing in a specific register only. Usually the terms chest register, middle register, head register, and falsetto register (or voice) were used. The singer was entirely free to use his/her own interpretation of these terms, but should strictly keep the intended register with no mixing of other registers. The mezzo-soprano singer made a distinction between chest voice, mid voice, and head voice (or falsetto). From the 1981 phonetogram recordings only graphical representations remained. In 2003, the phonetogram registration was repeated with the voice-profiler system by Pabon (www.voiceprofiler.com). In addition to fundamental frequency and vocal intensity, acoustic voice parameters were simultaneously recorded (Bloothooft & Pabon, 1999), out of which the crest factor has been used in this study. The crest factor describes the ratio of the maximum amplitude and the RMS value of a signal. Its value is maximal for a peaked signal and minimal for a sine wave. In this respect the crest factor relates to spectral properties and tends to indicate a flat spectrum for high values and a falling spectrum for lower values. Because of fatigue, the singer this time experienced difficulties to complete all the measurements for the full and separate register phonetograms. We will only present the result of all combined measurements.

## 3 Singer

The singer is a renowned Dutch mezzo-soprano, born in 1929. She had a long international career both in opera and in Lied singing. In 1981, at the age of 52, she was in her last years as a regular artistic performer. In 2003 she still worked as a vocal pedagogue[1].

## 4 $F_0$ decomposition

In general, the trace of the fundamental frequency of song can be thought to consist of three different components: (1) a base-line that follows the musical score, including effects of $F_0$ onset, $F_0$ transition between two tones, and over- or undershoot during this process, (2) a quasi-stationary, sinusoidal-shaped vibrato (modulation of $F_0$) with typical modulation frequencies between 4.5 and 9 Hz (Prame, 1995), and (3) fast irregular fluctuations, called jitter. Although each of these three components has its own specific frequency range, there can be some overlap which complicates the decomposition of the $F_0$ trace. Here we describe our method to accomplish the $F_0$ decomposition in the special case of vowels, individually sung at a prescribed $F_0$ frequency. Figure 1 shows a typical example of an original $F_0$ trace of a sung vowel, derived with Praat software (Boersma & Weenink, 1996), yielding fixed time samples, approximately corresponding to a waveform period[2].

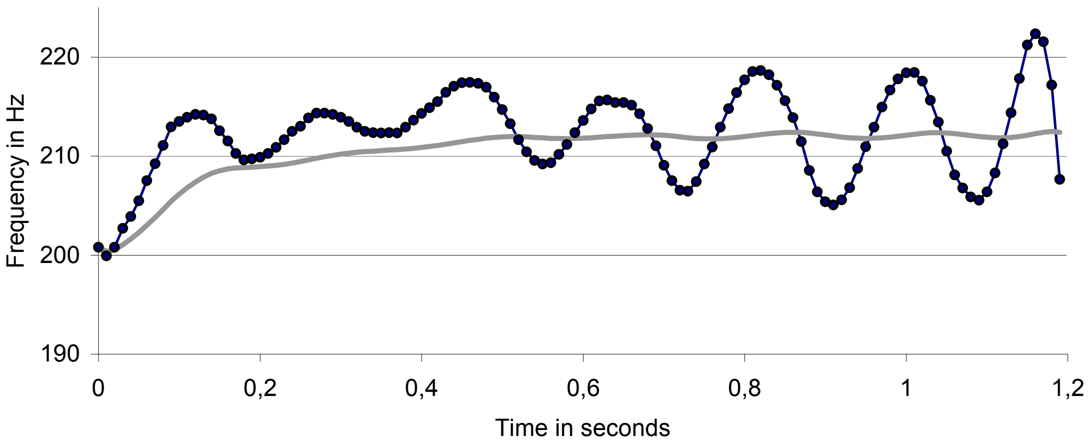


*Figure 1.* $F_0$ trace of the vowel /a/, sung by a mezzo-soprano in /hat/ context with an intended fundamental frequency of 220 Hz (dotted line). The grey line is the estimated transient component, produced by an adaptive low-pass filtering technique in which lowest cutoff frequency of 0.5 Hz is reached at 0.64 seconds.

The vowel onset often consists of a short time interval, between 100 and 200 ms long, in which $F_0$ starts at a low initial frequency (usually sung with low intensity) while increasing quickly towards the target value. During this onset the transient component 1 dominates. Usually within half a second from voice onset, vibrato (component 2) sets in. Although the singer will intend to keep the same average pitch, slow drift in $F_0$ may occur which should be considered as part of component 1. Finally, over the whole length of the $F_0$ trace small fluctuations may be observed (jitter, component 3, not so much present in professional singing). Computationally, we face a problem in the separation of the three $F_0$ components. If

---

[1] The singer suffered a stroke in 2000, but fortunately recovered well and retained full linguistic and vocal competencies. She could take up her pedagogical practice again.

[2] The Praat procedure includes an optimization procedure across a four-sample window as a result of which the $F_0$ values are less suited for the analysis of jitter, which was, however, no part of this study.

we separate vibrato by a fixed band-pass filter with cutoff frequencies of 2 Hz and 10 Hz, we will also include part of the $F_0$ transition during vowel onset. Moreover, vibrato usually is a quasi-sinusoidal modulation and includes higher frequency components, which will be lost in band-pass filtering. We therefore prefer to separate vibrato by removing the $F_0$ onset transition and subsequent possible drift by a dynamic low-pass filter with variable cutoff frequency. The latter is high during vowel onset — to capture the relatively rapid $F_0$ transient —but much lower soon thereafter — to capture average $F_0$ and slow $F_0$ drift. A running average procedure on the $F_0$ samples approximates this behavior. The grey line in Figure 1 shows the estimated transient and drift of $F_0$. The difference between both traces is an estimate of the true vibrato component, and is shown in Figure 2.

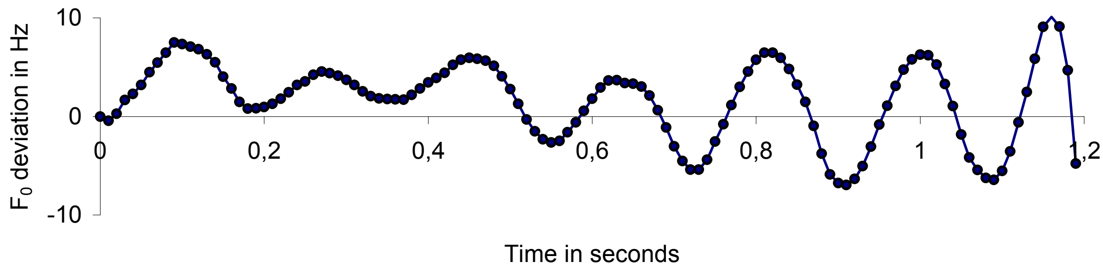


*Figure 2.* Remaining vibrato trace ($F_0$ deviation), after subtraction of the estimated $F_0$ transient, average, and drift from the original trace.

We would like to know not only the average vibrato frequency and vibrato extent (or modulation depth), but also the variation in both measures. For that, we need to compute an instantaneous estimate of the vibrato frequency and vibrato extent at all time samples. The method for this (an instantaneous frequency model) is described in the appendix. For the vibrato trace of figure 2, the resulting instantaneous vibrato frequency is shown in figure 3 and the instantaneous vibrato extent in Figure 4.

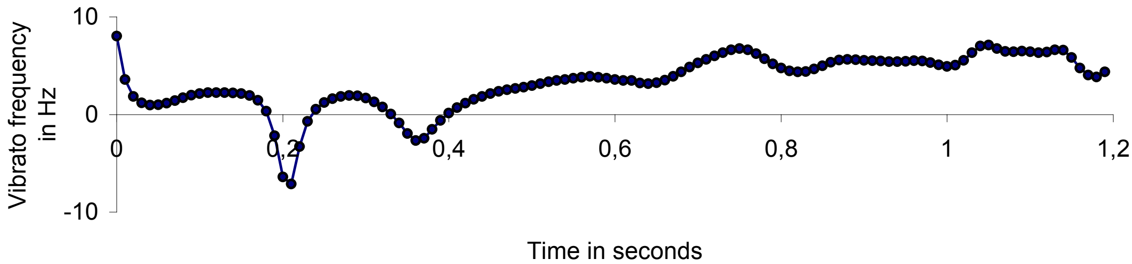


*Figure 3.* Instantaneous vibrato frequency of the trace of Figure 2.

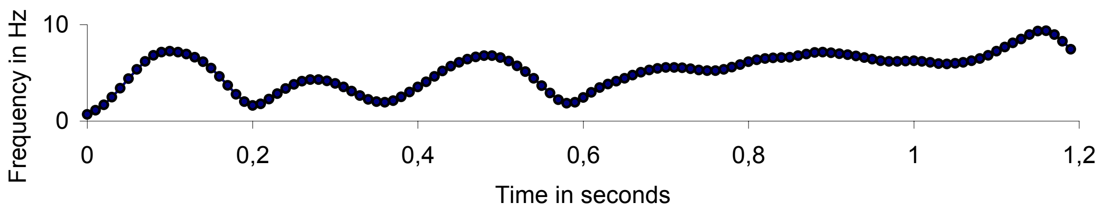


*Figure 4.* Instantaneous vibrato extent of the trace of Figure 2.

During the first 0.6 second of phonation, the instantaneous vibrato measures are quite unstable, and even have negative values. More study is needed to understand whether this is an intrinsic property of the voice or a computational artifact. Therefore, only data from two subsequent vibrato periods, selected from the second half of the vowels were used. For this interval we computed the average $F_0$ and the average and standard deviation of the instantaneous vibrato frequency. Instead of instantaneous vibrato extent, which is not further discussed in this paper, we used the difference in minimum and maximum $F_0$ during the chosen vibrato periods as a direct measure of vibrato modulation depth.

## 5 Results: phonetograms

In Figure 4 the phonetogram registration of 1981, at age 52 is shown. In the phonetogram the separate registrations of chest voice, mid voice and head voice are presented by vertical and horizontal hatching. The phonetogram at age 74 is presented in Figure 5 as an overlay of the one at age 52. Here, the grey-scale indicates the crest factor value: the darker, the higher the crest factor, which implies a more flat spectrum. When comparing both phonetograms, it is remarkable that the shape of the upper contour at high vocal intensity is quite similar (disregard the octave error at $F_0$ = 400 Hz in 1981), although there is a uniform loss in vocal intensity of about 6 dB at older age. At the high end of the frequency range, the singer has lost more than half an octave, while no losses are observed at low $F_0$. The differences at low vocal intensity may be due to the fact that the singer in 1981 probably sang at lowest acceptable singing level rather than lowest phonation level, which more likely is found between 55 and 60 dB SPL. In addition, measurement sensitivity at low vocal intensity was poorer in 1981.

The phonetogram from 1981 demonstrates that the chest voice ranged up to 300 Hz. This corresponds rather well to the highest crest factor levels seen in 2003, which up to 300 Hz mainly originate from the separate phonetogram recording in chest voice (not shown), and for higher $F_0$ values to a mixed register voice type.

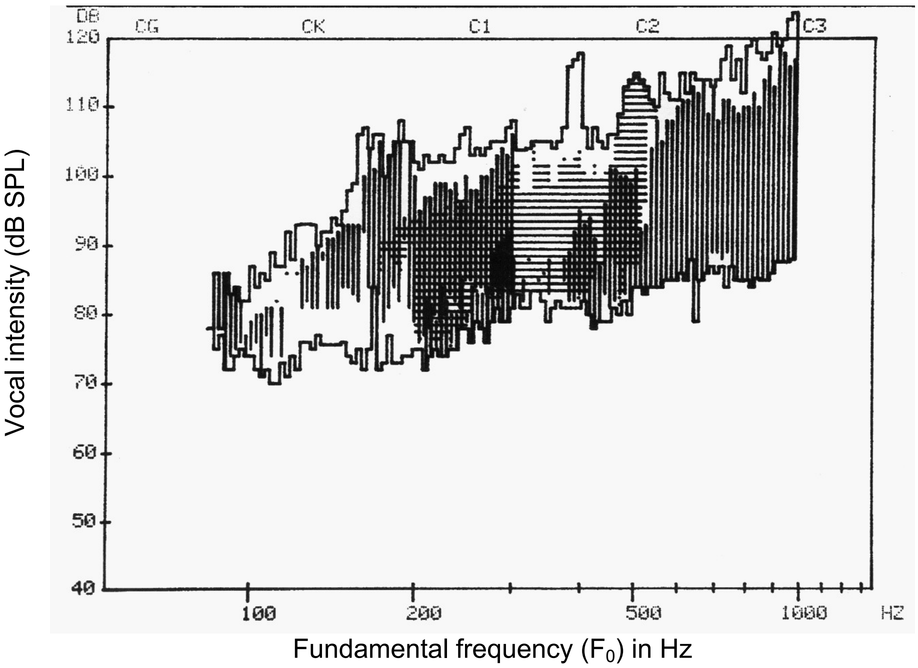


*Figure 4.* Phonetogram of the mezzo-soprano singer at age 52 (1981). Vertical and horizontal hatching indicates from left to right chest voice, mid voice and head voice.

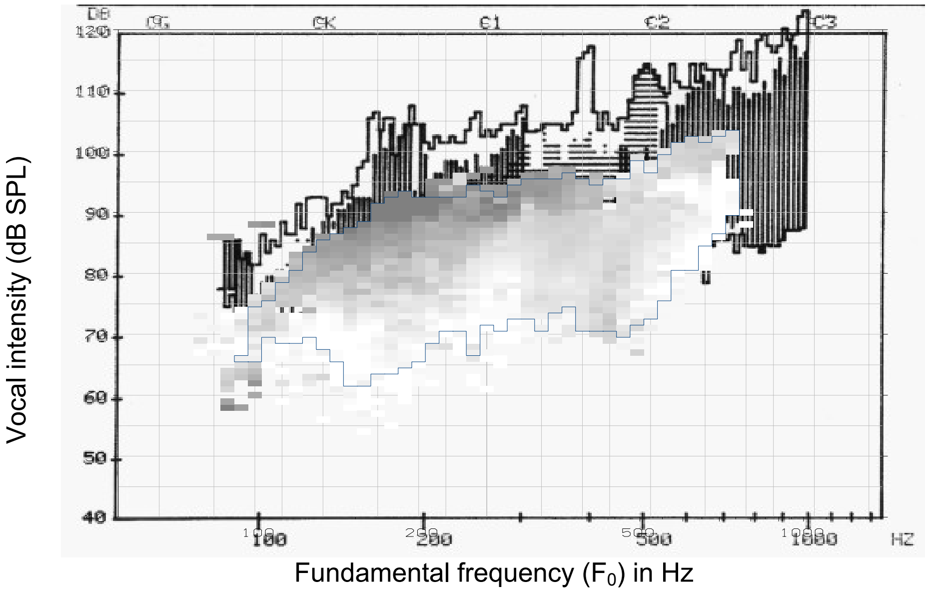


*Figure 5.* Phonetogram of the mezzo-soprano singer at age 74, as overlay of the phonetogram from 1981. The contour limits the area with more frequent phonations. The grey-scale indicates the level of the crest factor, which gradually changes from 3 dB (white) for a sinusoidal waveform, to 9 dB (dark) for a more peaked waveform.

The mid voice, as partly covering the higher chest voice and the lower head voice areas, roughly ranged from 200 to 600 Hz, which in 2003 still is a range with relatively high crest factors. The higher part of what the singer called head voice (falsetto) seems lost at later age.

## 6 Results: fundamental frequency and vibrato

At the fundamental frequencies of 220, 392 and 659 Hz, nine different vowels were each sung in five different modes, yielding 45 vowels at each pitch. Table 1 shows the target frequencies realized. The accuracy in reaching the target frequency was much better at age 52, although not really perfect. At later age, there was a serious undershoot, especially at higher fundamental frequencies. The singer noticed this herself during the recording of a vowel series. She also experienced some difficulties in register control during vowel onset at $F_0 = 659$ Hz. All this may relate to the fact that this frequency (E5 or e") was at age 74 at the top of the vocal range, while at age 52 it was in the centre of the head voice.

*Table 1.* Targeted and realized $F_0$ values in Hz, averaged over 45 vowels.

| Target $F_0$ | Realized $F_0$ | |
|---|---|---|
| | age 52 | age 74 |
| 220 | 211.5 ± 2.7 | 209.0 ± 4.5 |
| 392 | 385.0 ± 5.8 | 357.7 ± 9.1 |
| 659 | 632.2 ± 9.0 | 582.1 ± 26.7 |

In figure 6 we present a typical example of the distribution of the instantaneous vibrato frequency for a vowel /a/ sung at 392 Hz. The major difference between the distributions is exemplified by their standard deviation, which is 0.88 Hz at age 52 and 2.14 Hz at age 74, implying a more instable vibrato at later age. On average, across all vowels and $F_0$ values, the standard deviation of the instantaneous vibrato frequency was 1.19 Hz at age 52 and 3.73 Hz at age 74.

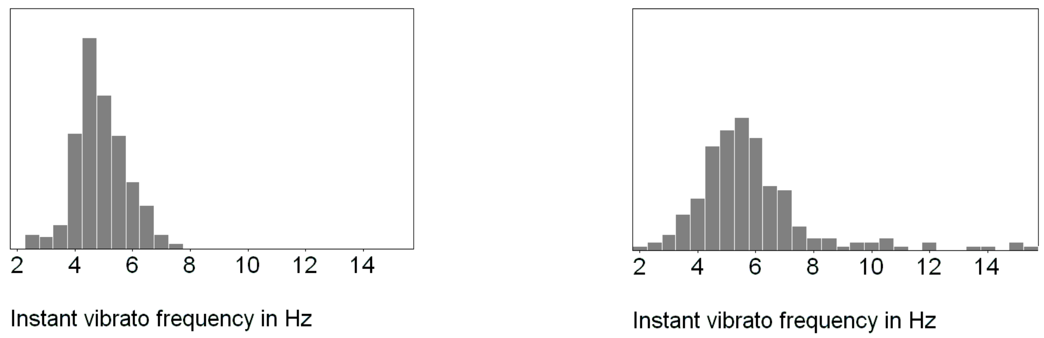


*Figure 6.* Normalized distributions of the instantaneous vibrato frequency for a typical vowel /a/, sung at 392 Hz in neutral mode. Left-hand panel at age 52 ($n$ = 322), right-hand panel at age 74 ($n$ = 205).

Not only is the stability of vibrato itself poorer at older age, also the variability in *average* instantaneous vibrato frequency per vowel is much larger. This is shown in the distributions presented in Figure 7. At age 52, the average vibrato frequency per vowel is well within a restricted range between 5 and 6.5 Hz (grand average 5.7 Hz), irrespective of the target $F_0$. At age 74, the distribution is much wider and even almost homogeneous at $F_0$ = 220 Hz, implying little control of vibrato, although the grand average (5.5 Hz) is not significantly different. Neither were any significant effects found of the five modes of singing.

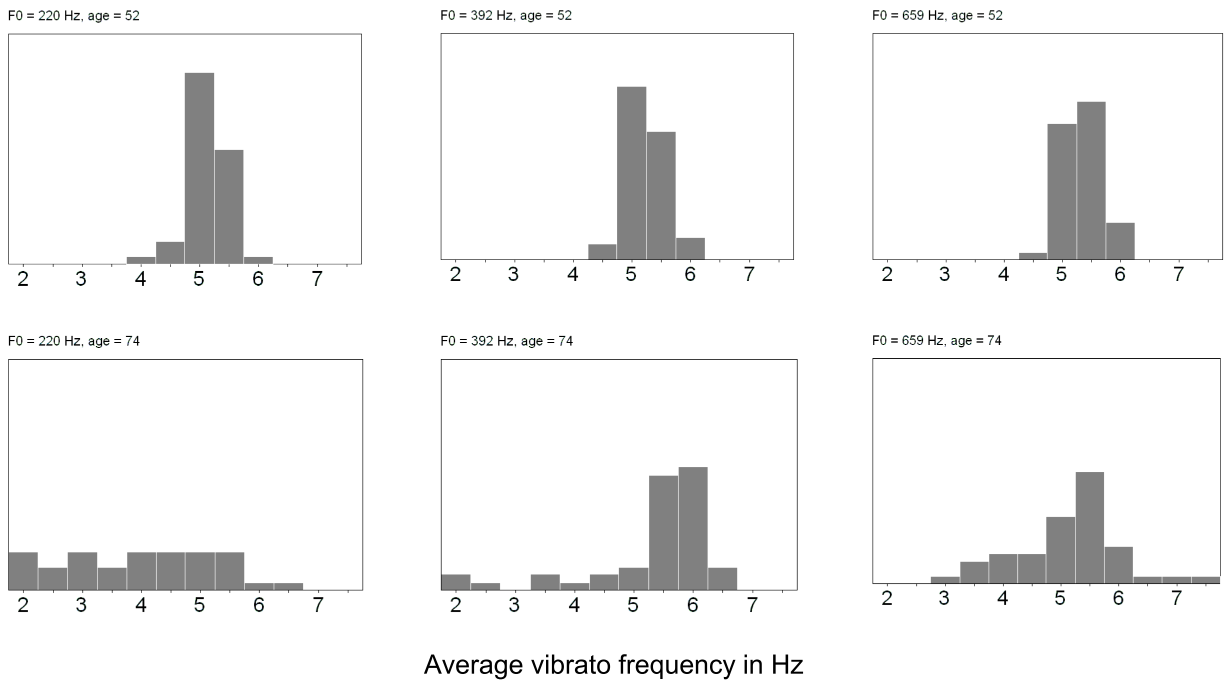


*Figure 7.* Distribution of the average vibrato frequency over all 45 vowels per age and per fundamental frequency.

Finally, we studied the modulation depth of the vibrato by taking the top-to-top frequency difference in $F_0$ during two vibrato periods. The result at $F_0$ = 220 Hz is presented in figure 8.

The mean at age 52 was 17.4 Hz; at age 72 this was reduced to 10.2 Hz. The same relative reduction was found for the other values of $F_0$. At age 52, these vibrato extent values correspond to those found by Seidner et al. (1995) for singers between 33 and 55 years of age, while those for our mezzo soprano at age 72 correspond in Seidner to values only found in (soft) singing with less expressive voice.

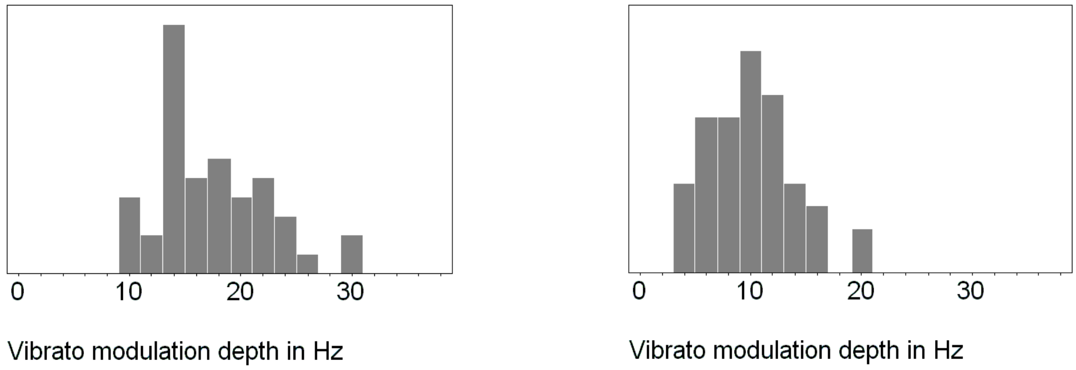


*Figure 8.* Distribution of the vibrato modulation depth (top-top) at $F_0$ = 220 Hz. Left-hand panel at age 52, right-hand panel at age 74, n = 45.

## 7 Discussion

A longitudinal analysis of singing is rare. This pilot study explored some possibilities for acoustic research on a mezzo-soprano singer for which comparable recordings were available at the age of 52 and 74. The literature shows that at least for the normal speaking voice noticeable acoustic changes in the voice set in at these decades. We found: (1) a considerable loss in the highest fundamental frequency range, (2) a loss of 6 dB at the highest vocal intensities, (3) less accuracy in targeting of $F_0$, (4) no significant change in average vibrato frequency, but (5) much more instability in vibrato frequency and less vibrato modulation depth. This implies a serious loss of the voice range and of vocal control. It should be realized that the effects of ageing can be manifold and that intersubject variability should be expected to be much higher than in young healthy voices, while fewer professional singers will be available for further study than non-singing subjects. Maybe the study of the ageing singing voice is deemed to be exploratory and difficult to generalize. A future attempt to record most of the 14 singers studied in 1981, will help us to come to grips with general and individual factors in the ageing professional singing voice.

## Appendix: Derivation of the instantaneous vibrato frequency and amplitude

A special method has been developed that allows evaluation of the sinusoidal smoothness of vibrato on even fragments of a vibrato period. For this, we return to the circular base of sinusoidal motion, in which a time step corresponds to an angle increment. Change in the instantaneous sine frequency then corresponds to a change in the angle increment step. This is known as an instantaneous frequency model, in which the distance to the origin corresponds to the instantaneous amplitude value for every time sample. We will demonstrate the application to vibrato modeling using a synthetic $F_0$ trace with an average of 250 Hz, vibrato frequency increasing from 1 to 10 Hz, and vibrato modulation depth increasing from zero to ±1 Hz (Fig. A1).

A two-dimensional sinusoidal representation of this trace is realized by first computing a 90° phase-shifted version of the trace (grey line in figure A.1). Subsequently, the original $F_0$

values are presented as X co-ordinates and the values of the 90° phase-shifted version as co-ordinates on the (imaginary) Y-axis (see figure A.2). In the resulting graph, the time axis is not an independent variable, but it curls through this plane according to the X/Y pairs, starting and ending in the origin. The spacing between time points is small in the beginning, representing a low instantaneous vibrato frequency, while spacing (or vibrato frequency) increases over time. We now smooth the trajectory and compute the instantaneous vibrato frequency from the rate of change in spacing. The result is shown in figure A.3, in which the exponential increase in instantaneous vibrato frequency is clearly visible. The rather strong ripples on the curve indicate the limited precision in the discrete approximation of the various computational steps.

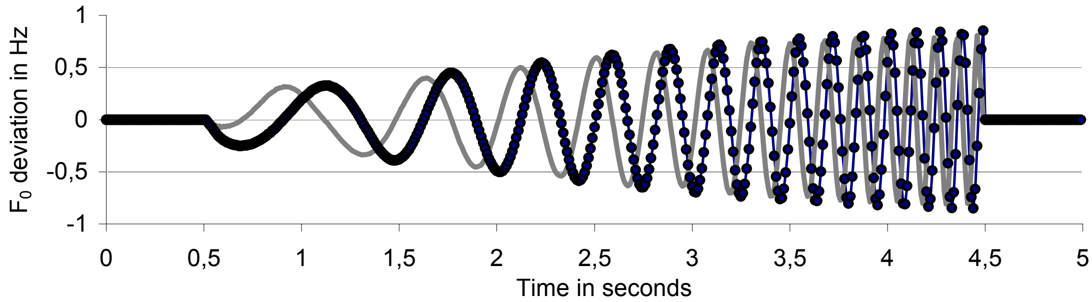


*Figure A.1.* Synthetic $F_0$ trace (fluctuations around an average of 250 Hz), with increasing modulation frequency and amplitude. The grey trace is the same signal but with a $90^{\circ}$ phase shift.

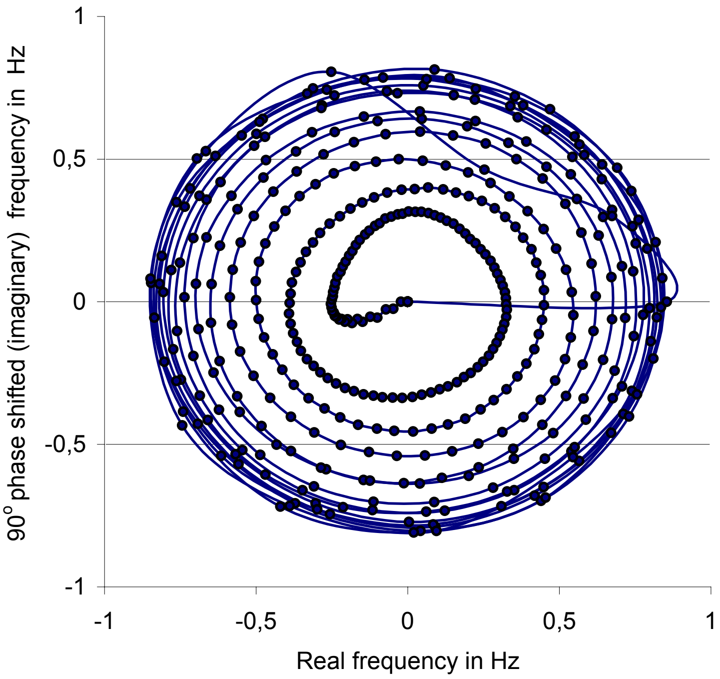


*Figure A.2.* Analytic signal plane with the values of the original $F_0$ trace along the X-axis and their 90° phase-shifted versions along the Y-axis, resulting in a circular trajectory.

The distance of the trajectory to the origin in Figure A.2 represents the instantaneous vibrato extent, which also increases gradually. The result is shown in Figure A.4. Notice that the given amplitude only has positive values and compares to half the top-to-top range of the vibrato modulations in the original $F_0$ trace. From the instantaneous vibrato and extent, a smoothed $F_0$ trace can be derived that approximates the original signal. Subtraction of the smoothed version from the original signal will reveal the short term differences, or jitter. In this example, this is not demonstrated.

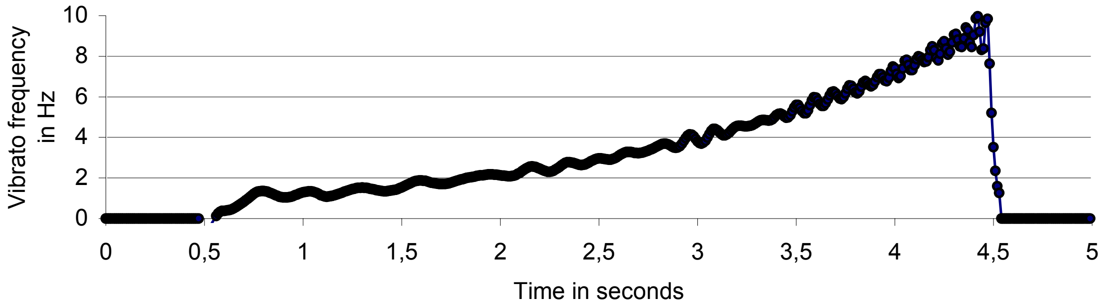


*Figure A.3.* Instantaneous vibrato frequency of the synthetic $F_0$ trace, derived from the smoothed instantaneous phase from the trajectory in Fig. A.2.

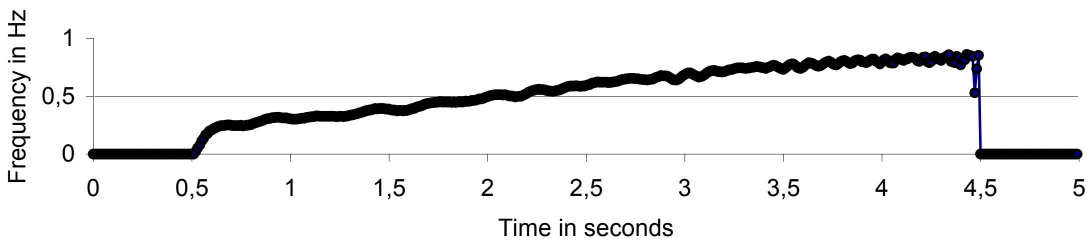


*Figure A.4.* Instantaneous vibrato extent of the synthetic $F_0$ trace.

## References

Bloothooft, G. (1981). A computer-controlled device for voice-profile registration. In *Proceedings IX Conference of the Union of European* Phoniatricians, Amsterdam (pp. 83-85).

Bloothooft, G. (1985). *Spectrum and timbre of sung vowels*. Doctoral issertation, Free University, Amsterdam.

Bloothooft, G., & Pabon, P. (1999). Vocal registers revisited. In *Proceedings Eurospeech '*99, Budapest (pp. 423-426).

Boersma, P., & Weenink. D. (1996). *Praat: Doing phonetics by computer*. Report nr. 132. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.

Böhme, G., & Hecker, G. (1970). Gerontologischen Untersuchungen über Stimmumfang und Sprechstimmlage. *Folia Phoniatrica*, *22*, 176-184.

Decoster, W. (1998). *Akoestische kenmerken van de ouder wordende stem.* Leuven: Leuven University Press.

Prame, E. (1995). Measurement of the Vibrato Rate of Ten Singers. In P.H. Dejonckere, M. Hirano, & J. Sundberg (Eds.), *Vibrato* (pp. 121-140). San Diego: Singular.

Seidner, W., Nawka, T. & Cebulla, M. (1995). Dependence of the Vibrato in Pitch, Musical Intensity, and Vowel in Different Voice Classes. In P.H. Dejonckere, M. Hirano, & J. Sundberg (Eds.), *Vibrato* (pp. 83-92). San Diego: Singular.