Pitch of synthetic sung vowels

Sundberg, J.

journal: STL-QPSR
volume: 13
number: 1
year: 1972
pages: 034-044

http://www.speech.kth.se/qpsr
B. PITCH OF SYNTHETIC SUNG VOWELS* 
J. Sundberg 

Abstract 
Experiments are reported in which musically trained subjects match the pitch of a stimulus tone with a variable response tone, both tones being synthesis of a sung vowel. The effects on pitch and pitch perception accuracy of vibratos of different kinds and the 'singing formant' are examined. The results appear to show that the pitch corresponds to the average fundamental frequency to the first approximation even though deviations from this average occur that may be explained by postulating a time constant in the fundamental frequency measuring system of the ear. Adding a vibrato or a 'singing formant' to a vowel sound affects neither the pitch perceived nor the accuracy of the pitch perception appreciably. The standard deviation values obtained from the experiments are considerably smaller than the difference limen for frequency probably owing to the subjects' musical training and to the fact that complex waveforms were used in the experiments.

Introduction 
In acoustic signals used for the communication of music, acoustic features can be recognized that appear to characterize a good musician. Probably such characteristics are not chosen randomly within a given cultural tradition, for music - as well as other arts - has a long history behind it and its development through the centuries is likely to obey to Darwinistic evolution principles: Such acoustic features can be expected to survive which are in some meaning adequate from the point of view of human perception. The fact that music can have a forceful impact on emotions can be interpreted as a support for this hypothesis. An important task of musical acoustics would therefore be to map such features and to examine whether their usage has a motivation in the way in which human perception works. Investigations of this kind may hopefully contribute to the understanding of our sensory and mental behavior.

The present paper reports an exploratory study aiming at an understanding of this kind of two acoustic features characterizing the type of professional singing that can be heard in our operas and concert halls. One feature is the vibrato and the other is the spectrum envelope peak near 3 kHz normally occurring in professional singing and frequently called the "singing formant". Their effects on pitch and pitch perception accuracy will be investigated.

The vibrato corresponds to a periodic low frequency modulation of the fundamental frequency. The waveform of this modulation is generally quite similar to a sinusoid and its periodicity varies between 5.5 and 8 cycles per second. The extent of the modulation is generally between ± 1.5% and ± 3% of the fundamental frequency, i.e. between ± a quartertone and ± a semitone (Seashore, 1932). In singing this modulation of the fundamental frequency is accompanied by a modulation of the overall sound pressure. Some authors have found that the two kinds of modulation are in phase, but others report that this is not always the case (Schoen, 1923; Sacerdote, 1957). The disagreement can probably be explained with reference to the relation between the strongest partial in the spectrum and the formant lying closest to this partial in frequency. If the formant frequency is higher than the frequency of the partial the amplitude and frequency modulations can be expected to be in phase whereas the opposite would occur when the strongest partial is slightly higher than its closest formant frequency. In any case the amplitude modulation, whether a consequence of the frequency modulation or not, is believed to be of minor importance as regards the sound perceived (Backus, 1969).

The vibrato has been shown to give a pleasing quality to the sound provided that it has a periodicity of slightly more than 6 Hz and a total extent of approximately ± 2% of the fundamental frequency. In this case one single pitch is perceived and even though as yet not strictly proved, this pitch is believed to correspond to the average of the fundamental frequency (Gibian, 1972; Ramsdell). It is also assumed that the pitch perceived from a vibrato tone is less certain than the pitch perceived from a steady state signal. It has been suggested that a function of the vibrato is to cover up slight errors in tuning, which, however, has not been brought to a strict scientific test (Stevens & Davis, 1938).

The "singing formant" has been studied previously mainly in male singing (Sundberg, 1970a and b; Sundberg, 1972). It appears to consist of a cluster of three formants. The lowest of these formants corresponds to the third formant in normal speech and the highest to the fourth. In-between these formants an extra formant is found. This formant can be ascribed to the larynx tube that may act as a separate Helmholtz resonator under certain conditions. These conditions involve a wide laryngeal ventricle and a wide pharynx. Additional conditions for generating a "singing formant" is to adopt
a special articulation of some vowels in singing as compared with speech. Thus, the "singing formant" is probably actively pursued by the singer at costs as regards his normal articulatory habits.

The perceptual function of the "singing formant" is not well understood. It has been shown that it adds to the "beauty" of the tone and that it may be correlated with a perceptual quality labeled "placement in the head" among singing teachers (Gibian, 1972). An interesting circumstance is also that the frequency of the "singing formant" approximately coincides with the region of maximal sensitivity of the human ear. Consequently the presence of the "singing formant" may make the sung sound easier to perceive, as suggested by Winckel (1971). We might expect that it also makes its pitch easier to discern.

Experimental procedure

The influence of the vibrato and the "singing formant" on pitch and pitch perception accuracy was studied by means of test series in which subjects matched the pitch of a given stimulus tone with a variable response tone. The stimulus and response tones were repeatedly presented binaurally to the subjects by means of ear phones. A pause of .5 sec separated the stimulus and the response tones, each of 1.9 sec duration. Between the response tone and the following stimulus tone the pause was 1.6 sec (Fig. III-B-1). The subjects adjusted the fundamental frequency of the response tone by turning a knob. When they found the pitch of the two signals identical they pressed a button intempting the signal presentation and triggering two frequency counters, one for each signal. In the case of vibrato tones the average fundamental frequency was measured.

When not otherwise stated the stimulus and response tones had the following characteristics. Both were synthesos of a baritone voice singing an [a]. The synthesis was performed on a formant synthesizer using the following formant frequencies: \( F_1 = 520 \text{ Hz}, F_2 = 955 \text{ Hz}, F_3 = 2385 \text{ Hz}, F_4 = 2550 \text{ Hz}, \) and \( F_5 = 3480 \text{ Hz}. \) The synthesis was found to be quite "natural" by all subjects. The sound pressure level in the ear phones varied around 70 ± 3 dB re. 0.0002 dyn/cm\(^2\) depending on the fundamental frequency. The rise and fall time of the stimulus and response tones was 80 msec. The vibrato, exclusively used for the stimulus tone, was generated by modulating the fundamental frequency sinusoidally at a rate of 6.5 cycles per sec and slowly
Fig. III-B-1. Oscillogram of the stimulus-response presentation (upper figure) and of the onset of the vibrato signal (lower figure).
growing to a maximal extent of $\pm 1.7\%$ of the fundamental frequency (Fig. III-B-1). As indicated above the "singing formant" was generated by connecting an extra formant circuit resonating at a frequency close to the third formant ($F_4 = 2550\) Hz). Four fundamental frequency values were used for the stimulus tone: 70, 115, 185, and 300 Hz. These frequencies are equally spaced over the range of a skilled professional bass singer. Each subject made 5 matchings of each of these fundamental frequencies in randomized order in one test series. Three or four such series, each lasting maximally 15 min, could be run in one test session without extending its duration over 60 min. In all sessions a rest of some minutes was offered to the subjects after each series.

The subjects were six musically trained persons in the age of 25 to 36 years. Three of them are professional musicians (MD, BT, NS) and all of them play or sing regularly in ensembles.

Experiment I: Test of method

Due to possible effects of auditory phenomena such as temporal masking etc. it cannot be taken for certain that subjects adjust the pitch of the response tone so that its fundamental frequency is identical with that of the stimulus tone in an experiment of the type described. Therefore, in one test series the stimulus and response tones were both presented without vibrato and with "singing formant", thus not differing from each other. 10 settings of each of the four fundamental frequencies were made by all subjects except MD and BT who made only 5. The results are plotted in Fig. III-B-2.a and b in terms of the individual subjects' means.

Fig. III-B-2.a shows that, on the average, there is no difference between the fundamental frequency of the stimulus tone $F_{0S}$ and that of the response tone $F_{0R}$. In some cases, however, individual subjects give values of $F_{0R}$ that differ from $F_{0S}$ by more than one standard deviation. Moreover the standard deviations vary considerably between subjects and fundamental frequencies between $0.05\%$ and $0.7\%$ (Fig. III-B-2.b). Therefore it seems motivated to judge the influence of a given stimulus property on the basis of the change in pitch and standard deviation that it imposes on the values of the individual subjects.

In the experiments to be described the stimulus and response tones were both generated by means of a formant synthesizer. Thus, the timbre and
EFFECT OF TEST SITUATION ON PITCH

\[ \frac{F_{0_r} - F_{0_s}}{F_{0_s}} \] (CENT)

ON PITCH

\[ \text{RISE} \uparrow\]

\[ \text{DROP} \downarrow\]

\[ \text{MEAN} \]

F0 (Hz)

ON PITCH ACCURACY

\[ \text{DECREASE} \uparrow\]

\[ s \text{(CENT)} \]

F0 (Hz)

Fig. III-B-2.a. Difference in fundamental frequency between the stimulus and response tones observed when they were generated in exactly the same way.

b. Corresponding values of the standard deviations. Symbols represent subjects (O = JS; ● = BN; △ = NS; ▲ = EJ; □ = BT; ■ = MD).
the sound intensity of both these signals depend on the fundamental frequency; every change in the fundamental frequency is accompanied by a number of shifts in the amplitudes of the individual partials depending on whether they approach a formant or not. Therefore the subjects may use similarity in timbre and loudness as additional cues for identity in pitch between the stimulus and response tones. Two test series were run in order to study if this was the case. In both series the stimulus tone included vibrato and "singing formant". In one test the response tone had the same characteristics as the stimulus tone except that it lacked vibrato. In the other test the response tone was a symmetrical squarewave signal thus considerably differing in spectrum from the stimulus tone (Fig. III-B-3). Moreover, its timbre was not affected by the fundamental frequency in the same way as the timbre of the stimulus tone. Four of the subjects participated in this test (JS, BN, NS, BT).

If other cues than identity in pitch were used by the subjects in the matching experiments, we would expect the standard deviation values to be higher when the response tone differed from the stimulus tone in the way just described. The standard deviation values from the two test series can be compared in Fig. III-B-4. The intersubject means lie very close together in the graph. Thus, the fact that identity in the timbre and intensity of the stimulus and response tones implied identity in fundamental frequency did not help the subjects in matching the pitch of the stimulus tone. Consequently, the standard deviations in the tests to be described are probably related to the accuracy with which the pitch is perceived.

Experiment II. Effect of vibrato

In this experiment the stimulus tone was presented with a vibrato modulating the fundamental frequency at a rate of 6.5 cycles per sec and to an extent of ± 1.7 %. Both stimulus and response tones included a "singing formant". The mean changes in the differences between the fundamental frequency of the response and the stimulus observed when the vibrato was added to the stimulus tone can be seen in Fig. III-B-5. Similarly, Fig. III-B-5.b gives the individual changes in the standard deviations. In this way Fig. III-B-5 illustrates the effect of adding a vibrato on pitch and pitch perception accuracy.
EFFECT OF SPECTRAL SIMILARITY STIMULUS-RESPONSE

ON PITCH ACCURACY

\[ s \text{(CENT)} \]

\[ F_0 \text{(Hz)} \]

--- ○ SIMILAR

□ DISSIMILAR

Fig. III-B-4. Standard deviation values observed when the stimulus and response were similar vs. dissimilar.
Fig. III-B-5. a. Change in the fundamental frequency difference between the stimulus and response tones observed after adding a sinusoidal vibrato to the stimulus tone.

b. Corresponding changes in the standard deviations. The subject symbols are the same as in Fig. III-B-2.
As regards pitch Fig. III-B-5. a demonstrates that the vibrato has a very small effect, if any. The individual differences ascribable to the vibrato do not reach even a 90% level of confidence in any case but one. The appropriate conclusion therefore seems to be that adding a vibrato has no effect on the pitch perceived.

As regards pitch perception accuracy the individual differences in the standard deviations given in Fig. III-B-5. b show that there might be a small effect when the fundamental frequency is as low as 70 or 115 Hz. For these frequencies the average standard deviation rises by .2 and .1% respectively, thus by an extremely small amount. The extent of the vibrato used in the experiments is rather small: ±1.7% of the fundamental frequency. Values of ±3% do occur in singing. One test series was run with one of the subjects in which the extent of the vibrato was increased to ±3%. However, no consistent changes in the standard deviations could be observed. Therefore the results can hardly be taken as a support for the hypothesis that a perceptual function of the vibrato is to cover up slight errors in tuning.

It is interesting that the pitch perceived of a vibrato tone appears to correspond to the averaged fundamental frequency. As long as a sinusoidal vibrato is used it cannot be decided whether this average is computed as the frequency midway between the extremes in the fundamental frequency modulation curve or whether it is the linear mean. If the modulation waveforms are distorted as shown in Fig. III-B-6 these two averages differ.

Two test series were run in which the stimulus had a vibrato obtained with the modulation waveforms shown in Fig. III-B-6. The rate of the vibrato was kept to 6.5 cycles per sec and the extent to ±1.7% of the fundamental frequency.

The individual mean values obtained from these test series can be compared with those from the test where both stimulus and response lacked a vibrato in Fig. III-B-7. a and b. This figure shows the effects of adding a vibrato of these kinds to the stimulus tone. As regards pitch there seems to be an effect of these distortions of the sinusoidal vibrato type. When the fundamental frequency makes positive deviations the response tone was adjusted slightly lower than the linear average, and when the vibrato waveform was inverted the opposite applies. Thus, the fundamental frequency of the response tone was in both cases deviating from the linear average in
Fig. III-B-6. Two waveforms used for the fundamental frequency modulation.
EFFECT OF VIBRATO TYPE

**ON PITCH**

![Graph](image)

**Delta F0R (CENT)**

**FO (Hz)**

**VIBRATO:**

- -

**RISE**

**DROP**

- -

**DECREASE**

**INCREASE**

**ON PITCH ACCURACY**

![Graph](image)

**Delta S (CENT)**

**FO (Hz)**

**VIBRATO:**

- -

**DECREASE**

**INCREASE**

Fig. III-B-7. a. Change in the difference in fundamental frequency between the stimulus and response tones observed after adding to the stimulus tone vibratos of the waveforms shown in Fig. III-B-4.

b. Corresponding changes in the standard deviations.
the direction of the steady state portion of the modulation waveform. This deviation may vary with the fundamental frequency since the mean curves separate more at higher frequencies. From Fig. III-B-7, b it can be seen that, on the average, there seems to be no effect on the standard deviation values. This indicates that the subjects did not find it more difficult to match the pitch of a stimulus with these types of vibrato than with a sinusoidal vibrato.

The results clearly show that the fundamental frequency determining the pitch of a vibrato tone is not the frequency midway between the extremes in the modulation curve. The linear average appears to be a better approximation, but the steady state portion of the modulation curve seems to be of somewhat greater importance than the rest of the curve.

Experiment III. Effect of the "singing formant"

In all experiments described above the stimulus and response tones contained a "singing formant". In one test series the stimulus and response tones were presented without this spectrum envelope peak. These signals were generated by simply omitting the extra formant at 2550 Hz in the synthesizer. The idealized spectrum envelopes are shown in Fig. III-B-8. The stimulus tone had a sinusoidal vibrato. The effect of adding a "singing formant" can therefore be studied by comparing the data collected in this experiment with the data obtained when the stimulus and response signals contained a "singing formant". Fig. III-B-9, a and b shows such a comparison.

The intersubject mean curve indicates that there might be an effect on the pitch perceived and that this effect depends on the fundamental frequency. However, statistically significant differences were found only in two cases (out of 24). Consequently the correct conclusion seems to be that the "singing formant" does not affect the pitch. Similarly Fig. III-B-9, b indicates that it does not affect the accuracy with which the pitch is perceived.

As previously mentioned it has been suggested that the "singing formant" makes the sound easier to perceive in large halls. Therefore it may be assumed that an effect could be observed if the stimulus tone is presented at a considerably lower sound pressure level than the response tone. In this case the "singing formant" would possibly facilitate the pitch perception. This hypothesis was tested in a test. The stimulus tone was presented at a sound pressure level of 45 dB rel. 0.0002 dyn/cm$^2$, i.e. 25 dB below the level of
Fig. III-B-8. Idealized spectral envelopes used for the stimulus and response tones in studying the effects of the "singing formant".
Fig. III-B-9. a. Change in the fundamental frequency difference between the stimulus and response tones observed after adding a "singing formant" to the stimulus tone.

b. Corresponding changes in the standard deviations. The subject symbols are the same as in Fig. III-B-2.
the response tone. In one test series the stimulus and response tones contained a "singing formant", in another not. The stimulus tone was presented with a sinusoidal vibrato. All subjects but one (EJ) participated in these series.

Fig. III-B-10.a and b shows the changes in the fundamental frequency of the response tone and in the standard deviations observed when a "singing formant" was added to the stimulus tone. Fig. III-B-10.a indicates that different subjects react very differently when the "singing formant" is introduced. At the same time there is no clear effect on the pitch accuracy. This is indicated by the very small changes in the standard deviation values as seen in Fig. III-B-10.b. Nevertheless significant changes in pitch occurs only spuriously. Thus it does not seem likely that the "singing formant" has any effect neither on pitch, nor on the accuracy with which the pitch is perceived even when added to a vowel of weak intensity.

In a previous investigation it was observed that the dependence on intensity of the pitch appeared to be different in complex sounds as compared with sinusoids (Lindqvist & Sundberg, 1971). Whereas the pitch of a sinusoid below about .5 kHz rises if the intensity is decreased, the pitch of a complex wave was found to drop with decreasing intensity. If this observation is applicable also to vowel sounds, subjects would match the pitch of the weaker sounds with a lower fundamental frequency than in the case when the stimulus and response tone were equally loud. Fig. III-B-11.a confirms that this is the case in terms of intersubject means of the fundamental frequency difference between the response tone and the stimulus tone. When the stimulus and response tones have the same intensity they differ very little in fundamental frequency. When the stimulus is 25 dB weaker than the response tone the pitch appears to drop considerably, no matter if a "singing formant" is present or not. It is interesting that this large shift in pitch is not accompanied by an increase in the accuracy of pitch perception, as indicated by the close agreement between the standard deviations given in Fig. III-B-11.b.

Discussion

According to the findings reported in this investigation the pitch perceived from a vibrato tone appears to correspond approximately to the average fundamental frequency. Moreover, the vibrato does not appear to affect the certainty with which the pitch is perceived, since the vibrato was not found
Fig. III-B-10. a. Change in the fundamental frequency difference between the stimulus and response tones observed after adding a "singing formant" to the stimulus tone. The sound pressure level of the stimulus tone was 25 dB below that of the response tone.

b. Corresponding changes in the standard deviations. The subject symbols are the same as in Fig. III-B-2.
Fig. III-B-11. a. Intersubject averaged difference between the stimulus and the response tones. The stimulus tone was presented at two intensities and at the lower intensity with and without a "singing formant".

b. Corresponding averages of the standard deviations.
to increase the standard deviations in the pitch matching experiments appreciably. Consequently, the vibrato can hardly reduce the demands on intonation accuracy in singing. These findings agree with the results of recent investigation in which the relations between fundamental frequency in singing and subjects' impression of bad intonation were examined: if the mean value of the fundamental frequency never coincides with the theoretically correct value during the course of a tone, musically trained subjects tend to judge the tone as "out of tune", no matter if the theoretically correct value lies within the frequency swing of the vibrato (Lindgren & Sundberg, 1972). On the other hand, the stimulus tones in our experiments certainly differ from the signals reaching the listeners' ears in a concert. For instance, one difference is that the average fundamental frequency of the stimulus tone remained constant (or, in the test with non-sinusoidal vibrato, almost constant) in the tests. This seems to be rare in musical performances. Transients in the average fundamental frequency may be of relevance to the effect of a vibrato as regards the pitch and the accuracy of pitch perception. Thus, a vibrato may be relevant when the singer's accuracy in tuning is uncertain. Another difference is that a voice is often accompanied by instruments or other voices in music. It is likely that a vibrato is capable of hiding beats between mistuned sounds under these conditions. Therefore, it may in such situations have a function of covering errors in tuning.

The average fundamental frequency used by the ear in pitch perception appears to be closer to the linear average than the frequency lying midway between the extremes in the modulation curve. On the other hand, the experiments with the non-sinusoidal vibratos also showed that the steady state portion of the fundamental frequency modulation waveform seems to be of slightly greater importance than the rapidly changing portions when the ear computes the average determining the pitch perceived. An interpretation of this cannot be made on the basis of the available set of data. However, it seems reasonable to assume that the explanation deals with some time constant in the fundamental frequency measuring system of the ear; this system might be too slow to follow the rapid changes in fundamental frequency occurring in the distorted modulation waveforms. This interpretation is in agreement with earlier findings where subjects matched the extremes in pitch between which a vibrato tone appeared to swing. These pitches correspond to frequencies slightly closer to the average frequency than the actual frequencies
of the extremes (Makepeace, 1939, cit. in Vennard, 1967). Further experiments with systematically varied shapes of the fundamental frequency modulation waveform may elucidate the properties of the fundamental frequency measuring system of the ear.

The "singing formant" seems to have no appreciable influence neither on the pitch nor on the accuracy of pitch perception. Thus no explanation to its presence in professional singing has been found. Another hypothetical explanation to be tested in future research is that it matches the average spectrum of an orchestral accompaniment in such a way that the tone sung is not masked.

The standard deviation values in all experiments described in this investigation are astonishingly small and appear to be rather unsensitive to variations of stimulus properties. The standard deviations in a test of this kind cannot be expected to be identical with the difference limen for frequency, but they could be expected to be of the same order of magnitude. However, the standard deviations are about 5 to 10 times smaller than the difference limen reported for sinusoids. One possibility of explaining this discrepancy is to assume that the difference limen for frequency reported in the literature is too large (Rakowski, 1972). It is also possible that the pitch of a complex sound is perceived with higher accuracy than the pitch of a sinusoid. In previous experiments musically trained subjects set the octave of given stimulus tones. In these experiments the standard deviations were approximately only doubled when the stimulus tone was shifted from a complex wave signal to a sinusoid (Lindqvist & Sundberg, 1971). Consequently it seems reasonable to assume that musical training is of appreciable importance to the difference limen for frequency. Most probably then the extremely small standard deviations found in the present test series are due to the musical training of the subjects and to the fact that the stimulus tone was a complex wave.

Conclusions

The pitch perceived from a vibrato tone appears to correspond to the average fundamental frequency to the first approximation. However, the measuring system of the ear is probably not capable of tracking very rapid changes in the fundamental frequency so that slight discrepancies may occur between the physical average and the average giving the pitch. The pitch perceived does not seem to be affected by the presence of a "singing formant", 
but the intensity may be of relevance. The pitch appears to drop when the intensity is reduced.

The accuracy of pitch perception does not seem to be influenced neither by a vibrato nor by a "singing formant". The standard deviations indicate that the difference limen for fundamental frequency is of the order of magnitude of .5 to .7 % for musically trained subjects matching the pitch of a complex sound.

Acknowledgments

The author is indebted to his colleagues J. Lindqvist and B. Lindblom for valuable discussions and to Special Instrument AB for providing their new gate system. The work was supported by the Bank of Sweden Tercentenary Fund under Contract No. 57/48.

References


Rakowski, A. (1972): "Accuracy of pitch measurements", paper given at the Psychoacoustics Symp. in Poznań, May-June 1972 by the chair of Acoustics at the University of Poznań, Poland.


