RECORDING THE OPERATIC VOICE FOR ACOUSTIC ANALYSIS

Densil Cabrera§, Pamela Davis #, Jennifer Barnes#, Margaret Jacobs* and David Bell#
§ School of Architecture, Design Science and Planning, # The Conservatorium of Music, and the * Faculty of Health Sciences, The University of Sydney NSW 2006.

ABSTRACT: This paper considers a number of factors related to the recording of the voices of operatic singers for acoustic analysis. We wanted to develop a technique for recording these singers in non-anechoic environments so we tested head-mounted microphones. We tested their effectiveness in recording the direct sound produced by the singer with relatively little interference from the reflections of the sound in the recording environment. We examined the effectiveness of various near-field microphone positions using a head and torso simulator in anechoic conditions and applied these observations to recording an operatic soprano, comparing the head-mounted and reference microphones. We also determined, that, for this singer, there was appreciable movement of the head and body during operatic singing, even when the singer tried to avoid moving.

1. INTRODUCTION

Successful operatic performances rely on the ability of singers to produce a voice that will be audible unamplified in large theatres, over an orchestral accompaniment and other singers. Audibility of an operatic voice in such circumstances appears to rely upon selective amplification of the voice in the part of the spectrum at which the human ear is most sensitive. The term "singer s formant" is in widespread use following observations that bass, baritone, tenor and mezzosoprano operatic singers project their voice over an orchestra by their 3rd, 4th and 5th formants with a relatively high energy peak at ~3 kHz [1-4]. The lower frequency range is important for articulatory and fundamental frequency features and there is harmonic energy up to at least 8 kHz. The recording system (microphone, pre-amplifier, recorder and digitisation hardware /software) must be able to encode the voice signal in this range, with as little error or distortion as possible and with an acceptable resolution so that features of the singing voice may be interpreted from spectral analyses.

There are other considerations in choosing a technique for the successful recording of operatic voices for research. One of these is that the singer may move during performance, depending on the expressivity of the music and the singer s emotional connection to it. Some opera pedagogues teach singers to use a shift of weight to the back foot just prior to a high note, which may be thought to assist postural support for the high note. Any shift of body weight or movement is likely to be associated with a change in the distance of the mouth to a fixed microphone. Mellody et al. [5] allowed operatic singers to move freely but this resulted in them reporting that the signal-to-noise ratio was affected by background noise, apparently from movement of their microphone which was mounted on the singer s clothing.

Singers are usually asked to maintain a constant microphone to mouth distance because the sound pressure level (SPL) diminishes as source-receiver distance increases, and a relatively small change in microphone to mouth position can significantly affect the measured SPL. Instructions to maintain a constant microphone to mouth distance, even by the placement of measuring rods between the microphone to rest on the subject s chin [6], may avoid these errors but vigilance is required on the part of the singer and the experimenter to monitor closely the distance during recording and to repeat tasks if movement occurs. It is likely that this level of vigilance on the part of the singer will interfere with his/her emotional expression, particularly in less experienced singers, and this may distract the singer from realising the experimental task successfully. Another approach has been to place a microphone on a harmonica holder on the subject s chest [7] although they acknowledge that head movements of the subject may result in microphone to mouth distance variation.

An alternative approach places the recording microphone at relatively longer distances from the singer, such as 40 cm [8], 50 cm [9] or 6 feet (approximately 183 cm) [10] as small variations of microphone to mouth distance will have a relatively smaller effect on the overall measured SPL. However, the transfer function from singer to microphone will be affected by reflections in the room, with the potential for significant comb-filtering effects. Furthermore, the longer-term room acoustical effects of reverberation and certain room resonance modes may also affect the measured sound spectrum. Closer recording distances will maximise the strength of the direct sound of the singer in non-anechoic rooms.

Head mounted microphones have been advocated recently for voice recording, with comparable data for voice features being recorded by a head-mounted condenser microphone as compared to a larger, professional grade, stand-mounted condenser microphone [11]. Microphone to mouth distance has been varied from 38 mm [12] to 51-76 mm [10]. A focus of this and other recent studies [11,13] has been how the choice of microphone may affect the measurement of parameters often used to describe impaired voice quality. There has been relatively little study on recording techniques appropriate for recording operatic voices.

There are other factors that may affect the selection of a head-mounted microphone technique to record operatic voices. Operatic singers may not always be available to be recorded in an anechoic chamber, or wish to sing in such an unusual artistic environment. A technique that would record the direct sound of the singer in preference to the reflected sound in the room would yield additional research recording opportunities if recording could take place in quiet rooms that were not sound-treated. A singer should be able to perform as naturally as possible, without limiting head and body movement, and so it is important to determine whether the mouth to microphone distance of a head-mounted microphone varies with performance and the effect of small variations in the recording distance on the recorded sound.

The vocal signal is not radiated uniformly around the head [14-16] and the sound spectrum is affected by the microphone position. The air flow from the mouth is likely to cause noise for a microphone in the air stream, making near-field axial microphone positions impractical without a substantial windscreen and high pass filtering. Dunn and Farnsworth [14] found that positioning a microphone 45¡ off axis produces little spectral weighting of speech (relative to the axial spectrum), and removes the air flow problem. For example a microphone 50 mm from the mouth, 45¡ to the side, suffers a 2 dB range of deviations from the axial spectrum for the eight 1/2-octave bands covering the 500 Hz — 8 kHz range.

Near-field mouth radiation data are also given by ITU-T Recommendations P.51 and P.58 [17-18], which specify the performance of an artificial mouth and a head and torso simulator. The mouth reference point of an artificial mouth is 25 mm in front of the lip plane on the mouth axis, and would be subject to air flow distortion in a real person. Like other published data, the ITU-T data show a relative reduction in high frequency sound pressure when the near-field measurement position is shifted off axis, with a greater reduction for larger angles.

A directional microphone, such as one with a cardioid directivity pattern, could be used in non-anechoic environments to reduce the effect of the indirect sound. However this approach has the accompanying problem of the proximity effect , where low frequencies are boosted in the near-field because of the microphone s sensitivity to pressure gradient [19]. A further practical disadvantage of a directional microphone is that its orientation must be carefully maintained.

As the measured pressure spectrum of a singer is affected by the microphone s relative position, comparing study results that use different microphone positions can be treacherous, especially as the ~3 kHz singer s formant is likely to be affected by microphone position. A utopian solution would be for all studies to apply the same measurement position. More practically, difference spectra between measurement positions could be used, tentatively, for comparisons between studies. However, the frequency-dependent radiation pattern varies with the mouth opening shape (and other factors), and hence between phonemes [15,16]. More research is required before such transformations can be executed with confidence using data from singers.

It is worth noting that in a typical unamplified operatic performance, the audience is likely to receive more reverberant sound energy originating from the singer than direct sound energy. This suggests that the overall sound power spectrum, integrating sound radiated to all directions, could make at least as good a reference measure as any of the pressure measurements previously discussed. Of course, the difficulty is measuring the sound power of a time-varying and moving performer — it is not really practical to do that in a naturalistic setting.

This aims of this paper are to determine: (a) the relationship between the sound power spectrum and pressure spectrum for various near-field microphone positions; (b) the optimal mouth to microphone distance and position for recording operatic voices using a commercially available head-boom mounted microphone; (c) whether a head-boom mounted microphone is successful in being able to record the direct sound produced by the singer with relatively little interference from the reflections of the sound in the recording environment; and (d) to report on applications of these techniques to record the singing voices of operatic singers.

2. SOUND POWER AND PRESSURE OF A HEAD AND TORSO SIMULATOR

Method

The Br el & Kj r Head and Torso Simulator (HATS) is equipped with an artificial mouth, consisting of a high compliance electrodynamic loudspeaker mounted inside the head, such that the sound is emitted from a rectangular mouth area 30 mm wide and 11 mm high. The mouth radiation properties comply with the cited ITU-T recommendations. The purpose of this part of the study was to relate a series of SPL measurements, close to side of the mouth, to the radiated sound power of the HATS. The interest in these side measurements was because it is quite easy to position a head-mounted microphone in line with the lips, horizontally displaced from the corner of the mouth.

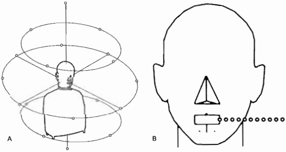

Figure 1. A. Illustration of the eighteen microphone positions used to determine the spatially integrated radiation spectrum and directivity pattern of the HATS. Each microphone position (represented by a small circle) was 1 m from the centre of the head at mouth height, and measurements were conducted in an anechoic room. B. Illustration of the eleven near-field microphone positions extending in a line to the left of the HATS mouth.

With the mouth radiating a constant pink noise stimulus, measurements were initially made in an anechoic room at a distance of 1 m from the centre of the head at mouth height, at 18 positions evenly distributed across a sphere surface (Fig. 1A). Eleven measurements were also made at 10 mm intervals in a line extending horizontally from the mouth edge to the side (Fig. 1B), and a twelfth at the mouth reference point. In the series of eleven measurements, the closest measurement was 0 mm from the mouth edge, meaning that the microphone was almost touching the HATS. A B&K 4135 1/4-inch microphone was used for all measurements described in this section. 1/3-octave measurements were made using a B&K 2131 spectrum analyser.

It is important to note that the axial distance measurements given in the ITU-T recommendations are from a lip plane 6 mm in front of the physical mouth orifice. However, the measurements reported in this paper are stated in terms of the distance from the physical mouth orifice. The principal reason for this is that we were interested in studying lateral microphone positions in line with the physical lip plane, because of the aforementioned simplicity in positioning microphones along this line.

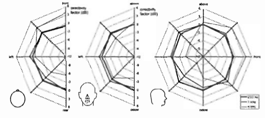

Figure 2. Octave band directivity factors for 250 Hz, 1 kHz and 4 kHz for the HATS at a distance of 1 m.

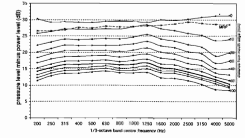

Figure 3. Difference between sound pressure level and sound power level for the eleven lateral measurement positions, as well as for the mouth reference point (MRP). The spectrum for 70 mm is in bold.

Results

In absolute terms, the measured sound power of the HATS is of little interest, because it depends on the signal spectrum and amplifier gain. However, as a reference, the sound power can be used in the determination of directivity factor, and also as a reference for the close microphone measurements. Directivity factor is the ratio of sound intensity radiated in a specific direction to the average intensity radiated in all directions, and can be expressed as a simple ratio, or else in decibels. Although measurements of directivity factors were made in 1/3-octave bands, they are presented just for three 1-octave bands in Figure 2 for the sake of succinctness and legibility. As might be expected, the directivity factors increase in the frontal axis in the high frequency range, due to the size of the mouth (as radiator) and head and torso (as barrier/reflector). The measurements also show that three of the 90¡ positions (left, right, and above) have directivity factors not far from 0 dB across the 200 Hz-5 kHz measurement frequency range, and might therefore be good choices for microphone positions if the overall radiated sound is of interest.

Similar results, for a custom-built dummy and for singers, can be found in Flanagan [15] and Marshall and Meyer [16] respectively, although those studies do not express the radiation in terms of directivity factor. The sudden decrease in high frequency intensity that occurs beyond 90¡ from the mouth axis is also found in those studies.

The results for the close measurements are shown in Figure 3. Between 30 mm and 100 mm the sound spectral components decrease by about 5 dB per doubling of distance, and this is reasonably independent of frequency. This may be contrasted with the 6 dB per doubling of distance (inverse-square law) which is found for this distance range in axial measurements [15, 18-19].

Figure 3 also shows the relationship between pressure level and power level at the mouth reference point (MRP), which is 26 mm in front of the physical mouth opening. As might be expected, the MRP sound pressure measurement contains relatively more high frequency sound than the power spectrum. On the other hand, most of the close lateral measurements show a reduction in high frequency sound (relative to the power spectrum), and a minor peak at 1.25 kHz.

The spectrum for the 70 mm distance is marked in bold because this distance was ultimately selected for measuring opera singers. The fact that the spectral contours in this region show a high degree of parallelism means that some inaccuracy in positioning a microphone will not significantly affect the resulting spectral profile.

3. TESTS OF A HEAD-MOUNTED MICROPHONE AND NON-ANECHOIC CONDITIONS

Method

Two miniature condenser omnidirectional microphones were tested for use in recording operatic voices: an AKG C477 WR oc/P (called here microphone 2) and an AKG 577 (called here microphone 3). The AKG C477 is manufactured with the microphone mounted on the left side and the AKG 577 is an earlier model not supplied with a head boom. These microphones were selected on the basis that the manufacturer s specifications include a relatively flat frequency response from 0.1-10 kHz, which encompasses the range of interest for operatic voice analysis, and a stated A-weighted signal-to-noise ratio of >68 dB.

The free-field frequency responses of these microphones were assessed by comparison to a B&K 4190 1/2-inch free-field microphone, using a JBL 4206 loudspeaker as the source, at a distance of 1 m on axis. The spectrum of a pink noise signal for each microphone in the reference position was obtained using a B&K 2034 FFT spectrum analyser. The frequency response of microphones 2 and 3 were determined by subtracting the sound level spectrum of the B&K 4190 from their raw sound level spectra.

Testing was carried out with author Jennifer Barnes, who is an operatic soprano rated as a "national — major principal [3.1a]" on the Bunch and Chapman [20] taxonomy of singers. Initial tests revealed that peak clipping was difficult to prevent with the AKG microphones 2 and 3 set at distances of 50 or 70 mm from the corner of the subject s lips, especially on the high soprano notes. The manufacturer s specifications are that the maximum SPL of the AKG C477 microphone is 133 dB (re 20 micropascals) and this seemed adequate for the very powerful operatic voice. A microphone pre-amplifier which permitted a two stage preamplification (Behringer 2200) solved the clipping problem, using auditory as well as visual monitoring via oscillographic real-time display of the recorded signal.

Comparison was made of simultaneously recorded AKG microphone 3 set at 50 mm laterally and a calibrated B&K 4190 microphone set at 300 mm on axis. The singer was recorded in a recording studio singing the first verse of *"Advance Australia Fair"*. Spectra of the recorded samples were digitised (sr=16 kHz, Loughborough Sound Images PC/C32 board) and analyzed using Soundswell Version 4.00 software (Hitech, Sweden). The two song raw spectra levels (125 Hz intervals) were calibrated to the SPL for the two calibration signals recorded for each microphone system. The spectrum recorded from AKG microphone 3 was corrected for its frequency response.

To assess movement of the microphone to mouth distance, the subject was fitted with microphone 2 on the head boom at 70 mm and asked to sing parts of an operatic song (*"Torna a Surriento"*) and an aria (*"Un bel di"* from *Madama Butterfly*) moving her head as little as possible. This was repeated with the AKG microphones set on fixed rods at various distances up to 100 mm from the side of her lips. A video camera was used to film her head and neck from the front and the side during singing. A 20 mm measurement grid was placed on a fixed rod at the level of the microphone and was used to measure the distance of the subject s lips to the boom-mounted microphone. Images of the singer and the microphones were printed and magnified, and distances were measured using the grid.

A software program called MLSSA (Maximum Length Sequence Signal Analyzer) was used, together with a sealed 4-inch loudspeaker (prior to the HATS s availability), to quantify the effect of the non-anechoic recording environment on the signals recorded from microphones 2 and 3. MLSSA uses a pseudo-random noise signal with a white spectral envelope to determine the impulse response of a system by cross-correlating the test signal with the measured signal. The impulse response shows the pattern of reflections produced by the system, as well as the general reverberant sound decay. The transfer function of the system is determined by applying a fast Fourier transform to the impulse response. Tests were done in a recording studio (10 m x 6 m plan, 0.4 s mid-frequency reverberation time) and a quiet office (3 m x 4 m).

Results

Frequency responses of the head-mounted microphones

Free field measurements found that the two AKG head-mounted microphones exhibit significant variation in sensitivity across the frequency range, with a notch in the vicinity of 3.5 kHz, and a peak around 4.0 kHz (Fig. 4). Microphone 2 had the wider variation, with sensitivity range of 8 dB between these frequencies. The large spectral irregularity in this region is of particular concern because it is in the range of the singer s formant. In fact, the singer s formant ranges between 2.5 and 3.5 kHz meaning that the microphone characteristics need to be factored into an analysis of this aspect of the singing voice.

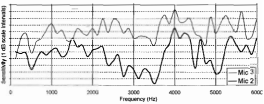

Figure 4. The measured free field frequency response of the two AKG head mounted microphones.

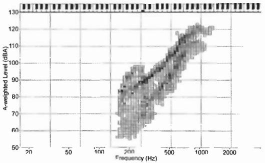

Figure 5. Phonetogram of a soprano voice, indicating a dynamic range of 63 dB.

Operatic voice recording

The phonetogram generated by the singer subject singing successive notes through her range is illustrated in Figure 5. Measured from microphone 2 at 70 mm lateral from the corner of the lips on the headphone boom, there was a 63 dB dynamic range over the singing range from 56 dB(A) on D3 (147 Hz) to 122 dB(A) on her top note E6 (1319 Hz). This is typical for soprano voices of similar experience and such a range presents a challenge to record the highest notes without distortion.

Extent and effect of head movement

During singing, the microphone to mouth (corner of lips) distance to the AKG microphones fixed on a rod placed lateral to the singer showed considerable change (range 8 mm shorter to 12 mm longer over various tasks) compared to the rest position. These observations were from video images recorded front-on and did not take into account the degree of front-back movement. However, two lateral video recordings revealed that there was considerable movement forward and backward with up to 80 mm movement between the lips and a fixed object during the operatic song and 100 mm for the more emotional aria. This lateral video recording also revealed much greater vertical movement (up to 80 mm) during performance of the aria. Such movements imply large changes in the mouth-microphone transfer function.

There were smaller changes in the microphone to mouth distance during singing using the head-mounted C477 (mean 4 mm longer, SD 2 mm) and some of these changes appeared to relate to changes in mouth shape as the mouth opened and closed for the various phonemes, a topic which warrants more detailed study. There appeared to be negligble front-back and up-down movement artefact as the microphone moved with the singer.

Effect of non-anechoic environments

The recording studio used to record the singer exhibited a number of room reflections (from the MLSSA data), the strongest being at 21.5 ms at —30 dB relative to the direct impulse. This corresponded to a glass wall 3.4 m in front of the subject. When the equivalent FFT plots for the anechoic room and the recording studio were overlaid, there was negligible effect in the frequency range of interest, with less than 1 dB difference in the high range. In the office, two reflections were apparent around 12 ms and a level of —30 dB, corresponding to a large glass window. When the anechoic and office data were compared, there was less than 2 dB differences across the spectrum range of interest. No voice recordings were carried out in this office although this was used for subsequent experiments recording opera singers simultaneously with laryngeal endoscopy [21-22].

4. DISCUSSION

This paper has described a technique to record operatic voices with a wide dynamic range. A head-mounted microphone allows the microphone position relative to the mouth to be well-defined and stable. Positioning the microphone to the side avoids the problem of air-flow-induced noise, and also has the advantage of a simple measurement for positioning. The close side position suffers a loss of high frequency sensitivity, compared both to the axial frequency response and the overall sound power frequency response.

The AKG C477 and C577 head-mounted microphones show a degree of spectral irregularity which would substantially influence the interpretation of data (such as formant frequencies and strengths) unless compensation is applied. Measurement grade microphones, which would not require compensation, have rarely been used in singing studies.

Previous studies have used an anechoic chamber to achieve accuracy in recording [1,3]. The recording system described here, which optimises the singer s sound in preference to the recording environment, has been particularly useful in recruiting professional operatic singers who have volunteered to be recorded in places such as a rehearsal room near their workplace rather than having to travel to an anechoic chamber with its associated artistic unfamiliarity.

Movements of a singer relative to a fixed microphone seems unavoidable even when the singer is unusually disciplined (as our subject was) and attempting to remain as still as possible. These variations in the mouth to microphone distances appear to preclude the use of fixed microphones in the near field for recording voice for acoustic analysis.

A recent paper Mellody et al. [5] recorded operatic voices using a body-mounted microphone and reported that the recording levels were continuously altered during recording to avoid overloading due to the dynamic range produced by the singers. This technique of optimising the energy levels is common among audio engineers in music theatre, television and radio broadcasts but it effectively precludes meaningful acoustic analysis of the voice energy. In another recent study of soprano vowels, Weiss et al [10] reported that strong voices that were close miked (two to three inches from the mouth) produced an "almost square wave pattern" which assumed a more normal shape as levels decreased, suggesting that the recording levels were too high or, more likely, the microphone had been overloaded at these high levels of operatic voices. These reports add to some of our earlier observations that the dynamic range of operatic voices recorded with close microphones deserves special consideration.

The well-known observation of the directivity of the higher frequencies, illustrated in Figure 2, has important pedagogical applications. A singer hears their singing voice with the air-conducted sound having a considerable loss of energy in the "singer s formant" region. It is understandable why young singers need to rely upon recordings made from the front and upon their pedagogue s ear. The successful opera singer appears to have the ability to rely upon the proprioceptive sensations from their body to achieve projection over an orchestra, more so than their auditory feedback. This must be a very hard lesson for young opera singers as it is this quality that translates to audition success. This area warrants further investigation, as it has practical application to pedagogy.

REFERENCES

1. G. Bloothooft, and R. Plomp, "The sound level of the singer s formant in professional singing" *J. Acoust. Soc. Am* **79**, 2028-33 (1986)
2. J. Sundberg, "Articulatory interpretation of the singing formant" *J. Acoust Soc. Am.* **55**, 838-844 (1974)
3. J. Sundberg, " The Acoustics of the Singing Voice" *Scientific Am.* **3**, 82-89 (1977)
4. J.Sundberg, "Level and center frequency of the singer's formant" *J. Voice* **15**, 176-86 (2001)
5. M. Mellody, F. Herseth, and G.H. Wakefield, "Modal distribution analysis, synthesis, and perception of a soprano's sung vowels" *J. Voice* **15**, 469-482 (2001)
6. J. Michel and J. Grashel, "Vocal vibrato as a function of frequency and intensity", In: Lawrence V.L., ed., Transcripts of the Ninth Symposium: Care of the Professional Voice, New York, The Voice Foundation, 45-8 (1980).
7. D.B. Price & R.T. Sataloff, "Technical note: a simple technique for consistent microphone placement in voice recording" *J. Voice* **2**, 206-207 (1988).
8. S.D. Foulds-Elliott, P.J. Davis, S.J. Cala, and C.W. Thorpe, "Respiratory function in operatic singing: effects of emotional connection" *Logopedics*

Phoniatrics Vocology 25: 152-169 (2000)
9. J.P.H. Pabon, and R. Plomp, "Automatic phonetogram recording supplemented with acoustical voice-quality parameters" *J. Speech Hear. Res.* **31**, 710-722 (1988)
10. R. Weiss, W.S. Brown, Jr. and J. Morris, "Singer's formant in sopranos: fact or fiction?" *J. Voice* **15**, 457-68 (2001)
11. W.S. Winholtz, and I.R. Titze, "Miniature head-mounted microphone for voice perturbation analysis" *J. Speech, Lang. Hear. Res.* **40**, 894-899 (1997)
12. R.E. Stone, C.J. Bell, and T.D. Clack, "Minimum intensity of voice at selected levels within the pitch range" *Folia Phoniatrica* **30**, 113-118 (1978)
13. V. Parsa, D.G. Jamieson and B.R. Pretty, "Effects of microphone type on acoustic measures of voice" *J. Voice* **15**, 331-343 (2001)
14. H.K. Dunn, and D.W. Farnsworth, "Exploration of pressure field (sic) around the human head during speech" *J. Acoust Soc. Am.* **10**, 184-199 (1939)
15. J.L. Flanagan, "Analogue measurements of sound radiation from the mouth" *J. Acoust Soc. Am.* **32**, 1613-1620 (1960)
16. A.H. Marshall and J. Meyer, "The directivity and auditory impression of singers" *Acustica* **58**: 130-140 (1985)
17. International Telecommunication Union (1996) ITU-T Rec. P.51 Artificial mouth.
18. International Telecommunication Union (1996) ITU-T Rec. P.58 Head and torso simulator for telephonometry.
19. J. Eargle, *The Microphone Handbook.* ELAR Publishing, Commack NY. (1981)
20. M. Bunch, and J. Chapman, "Taxonomy of Singers Used as Subjects in Research" *J. Voice* **14**, 363-369 (2000)
21. P. Davis, J. Oates, D. Kenny, J. Livesey, J. Chapman, and D. Carbrera, "Singer s formant characteristics in a group of opera chorus artists" *The Australian Voice Association 6th Voice Symposium, Adelaide,* October (2002) http://www.plevin.on.net/voice2002/
22. M. Jacobs, P. Davis, J. Livesey, and A. Connolly "Acoustics and perception of vocal control in operatic sopranos following anaesthetic and endoscopy" *The Australian Voice Association 6th Voice Symposium, Adelaide*, October (2002) http://www.plevin.on.net/voice2002/