# TMH-QPSR Abstracts

## (Quarterly Progress and Status Report)

Now only on line: http://www.speech.kth.se/qpsr/tmh/

## SPEECH

### Measuring the effectiveness of word prediction: The advantage of long-term use

*Tina Magnuson and Sheri Hunnicutt*

This paper provides results regarding keystroke savings and text composition time from a long-term evaluation of a Swedish word prediction program. The evaluation was carried out as a single-case pilot study for one participant with dyslexia and no motor disabilities. During 13 months of training, during which a speech synthesis facility was introduced, his performance was shown to improve both with and without the word prediction program after an initial increase in text composition time. On average, he was able to save 22.7% total keystrokes with word prediction only and 30.5% with word prediction combined with speech synthesis. Text composition time increased upon introduction of the word prediction program by at least 22 %. Over time, however, and with the further introduction of speech synthesis, he was eventually able to return to his original text entry speed.

*TMH-QPSR Vol 43/2002: 57-67*

## MUSIC

### Choir acoustics – an overview of scientific research published to date

*Sten Ternström*

Choir acoustics is but one facet of choir-related research, yet it is one of the most tangible. Several aspects of sound can be measured objectively, and such results can be related to known properties of voices, rooms, ears and musical scores. What follows is essentially an update of the literature overview in my Ph.D. dissertation from 1989 of empirical investigations known to me that deal specifically with the acoustics of choirs, vocal groups, or choir singers. This compilation of sources is no doubt incomplete in certain respects; nevertheless, it will hopefully prove to be useful for researchers and others interested in choir acoustics.

*TMH-QPSR Vol 43/2002: 1-8*

### Effects of a velopharyngeal opening on the sound transfer characteristics of the vowel [a]

*Birch P, Gümoes B, Prytz S, Karle A, Stavad H, and Sundberg J*

Many professional operatic singers sing the vowel [a] with a velopharyngeal opening (VPO) (Birch & al, in press). Here we analyse the resonatory effects of a VPO. On the basis of CAT scan imaging of a singer's vocal and nasal tracts acoustic models were constructed in which VPOs were modelled by different

tubes. The sound transfer characteristics of this model were determined by means of sine-tone sweep measurements. The VPO introduced a zero in the frequency curve of the model at the nasal cavity resonance. This zero can presumably be tuned to the first formant of an [a] which would attenuate the first formant, such that the higher formants become more dominating in the spectrum.

*TMH-QPSR Vol 43/2002: 9-15*

## Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in a professional female singer

*Stone RE (Ed), Cleveland TF, Sundberg J, and Prokop J*

Understanding how the voice is used in different styles of singing is commonly based on intuitive descriptions offered by performers who are proficient in only one style. Such descriptions are debatable, lack reproducibility and lack scientifically derived explanations of the characteristics. We undertook acoustic and aerodynamic analyses of a female subject with professional experience in both operatic and Broadway styles of singing, who sang examples in these two styles. How representative the examples are of the respective styles was investigated by means of a listening test. Further, as a reference point, we compared the styles with her speech. Variation in styles associated with pitch and vocal loudness was investigated for various parameters: subglottal pressure, closed quotient, glottal leakage, $H_1$-$H_2$ difference (the level difference between the two lowest partials of the source spectrum), and glottal compliance (the ratio between the air volume displaced in a glottal pulse and the subglottal pressure). Formant frequencies, long-term-average spectrum and vibrato characteristics were also studied. Characteristics of operatic style emerge as distinctly different from Broadway style, the latter being more similar to speaking.

*TMH-QPSR Vol 4/2002: 17-29*

## Spectral distribution of solo voice and accompaniment in pop music

*Daniel Zangger Borch and Johan Sundberg*

Singers performing in popular styles of music mostly rely on feedback provided by monitor loudspeakers on the stage. The highest sound level that these loudspeakers can provide without feedback noise is often too low to be heard over the ambient sound level on the stage. Long-term-average spectra of some orchestral accompaniments typically used in pop music are compared with those of classical symphonic orchestras. In loud pop accompaniment, the sound level difference between 0.5 and 2.5 kHz is similar to that of a Wagner orchestra. Long-term-average spectra of pop singers' voices showed no signs of a singer's formant but a peak near 3.5 kHz. It is suggested that pop singers' difficulties to hear their own voices may be reduced if the frequency range 3-4 kHz is boosted in the monitor sound.

*TMH-QPSR Vol 4/2002: 31-35*

## Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos

*José Bretos and Johan Sundberg*

Two high pitched and long sustained notes, F5 and A5, were selected from an aria for soprano in ten commercial recordings of G Verdi's opera Aida. Both notes are sung without any instrumental accompaniment and with a crescendo. These audio examples were analyzed with regard to fundamental frequency, vibrato rate, vibrato extent, intonation and sound level and the relationship between these parameters was examined. The results reveal that vibrato rate differed significantly between the tones in most of the singers and confirm Prame's observations that vibrato rate tends to increase exponentially toward the end of tones. Moreover, both vibrato extent and mean F0 often varied systematically with sound level. The regularity of the vibrato tended to be greater at F5 than at A5.

*TMH-QPSR Vol 4/2002: 37-44*

## A flow waveform adaptive mechanical glottal model

*Carlo Drioli*

A waveform adaptive physical model of the glottal source is proposed. The model relies on a lumped mechano aerodynamic schema loosely inspired to the one- and two-mass lumped models. The vocal folds are represented by a single mechanical resonator and a propagation line which takes into account the vertical phase differences. The vocal folds displacement is coupled to the glottal flow by means of an aerodynamic driving block which includes a general parametric nonlinear component. The principal characteristics of the flow-induced oscillations are retained, and the overall model is able to adapt to glottal flow signals with different characteristics.

*TMH-QPSR Vol 4/2002: 69-80*

## Effects of a professional solo singer education on auditory and kinesthetic feedback - a longitudinal study of singers' pitch control

*Dirk Mürbe, Friedemann Pabst, Gert Hofmann, and Johan Sundberg*

The significance of auditory and kinesthetic feedback to pitch control in singing was described in a previous report of this project for students at the beginning of their professional solo singer education (Mürbe et al., 2002). Since it seems reasonable to assume that pitch control can be improved by training, the same students were reinvestigated after 3 years of professional training. As in the previous study, the singers sang an ascending and descending triad pattern with and without masking noise in legato and staccato and in a slow and a fast tempo. Fundamental frequency and interval sizes between adjacent tones were determined and compared to their equivalents in the equally tempered tuning. The average deviations from these values were used as estimates of intonation accuracy. Intonation accuracy was reduced by masking noise, by staccato as opposed to legato singing and by fast as opposed to slow performance. The contribution of the auditory feedback to pitch control was not significantly improved after education while the kinesthetic feedback circuit was improved in slow legato and slow staccato tasks. The results support the assumption that the kinesthetic feedback contributes substantially to intonation accuracy.

*TMH-QPSR Vol 4/2002: 81-87*

## Estimating perceived phonatory pressedness in singing from flow glottograms

*Johan Sundberg, Margareta Thalén, Paavo Alku, and Erkki Vilkman*

The normalized amplitude quotient (NAQ), defined as the ratio between the peak-to-peak amplitude of the flow pulse and the negative peak amplitude of the differentiated flow glottogram and normalized with respect to period time, has been shown to be related to glottal adduction. Glottal adduction, in turn, affects mode of phonation and hence perceived phonatory pressedness. The relationship between NAQ and perceived phonatory pressedness was analyzed in a material collected from a professional female singer and singing teacher who sang a triad pattern in breathy, flow, neutral and pressed phonation in three different loudness conditions (soft, middle, loud). In addition, she also sang the same triad pattern in four different styles of singing, classical, pop, jazz and blues, in the same three loudness conditions. A panel of nine experts rated the degree of perceived phonatory press along visual analogue scales. Comparing the obtained mean rated pressedness ratings with the mean NAQ values for the various triads showed that about 73% of the variation in perceived pressedness could be accounted for by variations of NAQ.

*TMH-QPSR Vol 4/2002: 89-96*

# HEARING

## Automatic classification of the telephone listening environment in a hearing aid

*Peter Nordqvist and Arne Leijon*

An algorithm is developed for automatic classification of the telephone-listening environment in a hearing instrument. The system would enable the hearing aid to automatically change its behavior when it is used for a telephone conversation (e.g., decrease the amplification in the hearing aid, or adapt the feedback suppression algorithm for reflections from the telephone handset). Two listening environments are included in the classifier. The first is a telephone conversation in quiet or in traffic noise and the second is a face-to-face conversation in quiet or in traffic. Each listening environment is modeled with two or three discrete Hidden Markov Models. The probabilities for the different listening environments are calculated with the forward algorithm for each frame of the input sound, and are compared with each other in order to detect the telephone-listening environment. The results indicate that the classifier can distinguish between the two listening environments used in the test material: telephone conversation and face-to-face conversation. [Work supported by GN ReSound.]

*TMH-QPSR Vol 4/2002: 45-49*

## The relationship between residual hearing and speech intelligibility - Is there a measure that could predict a prelingually profoundly deaf child's possibility to develop intelligible speech?

*Anne-Marie Öster*

The relationship between speech intelligibility of eleven prelingually deaf children and their residual hearing capabilities was investigated. Audio recordings were made of the deaf children, ranging from eleven to seventeen years of age, when they read so-called Helen-questions. Tapes containing the Helen-questions were played to inexperienced and experienced normally hearing listeners to measure the intelligibility of each child's speech. The speech intelligibility scores were related to the amount of residual hearing, measured as the better-ear average of pure-tone thresholds at 500, 1000 and 2000 Hz, shape of audiogram and the degree to which the children could use their residual hearing. The purpose was to investigate if any of these measures of residual hearing might be a good predictor of a prelingually deaf child's possibility to develop intelligible speech. The best correlation was found between the children's ability to use their residual hearing for speech perception and the intelligibility of their speech.

*TMH-QPSR Vol 4/2002:* 51-56