A virtual real time room simulation to determine the influence of room acoustics on voice quality

Philipp Heck[1], Malte Kob[1], Tobias Lentz[2], Christiane Neuschaefer-Rube[1]

[1] *Dep. of Phoniatrics, Pedaudiology, and Communication Disorders, RWTH Aachen University – University Hospital, Pauwelsstr. 30, D-52074 Aachen, Germany*

[2] *Inst. of Technical Acoustics, RWTH Aachen University, Templergraben 55, D-52072 Aachen, Germany*

Introduction

Professional speakers often suffer from voice diseases that are partially caused by uncomfortable conditions in the performance room. Under these conditions, reverberation time and background noise are of great importance. Whereas the interaction between room and voice has been extensively studied with focus upon perceptual questions such as room-listener interactions, the voice performance of professional speakers influenced by room conditions has not yet been thoroughly investigated. One reason might be that for the assessment of voice quality in the actual environment of a speaker, both subjective and objective methods would suffer from ambient noise as well as from alteration of the voice signal, due to reflections within the room. With a simulated acoustical environment, this disadvantage can be avoided. Another advantage of the examination by help of a virtual room simulation is the possibility to accomplish the tests with a reliable setup that provides reproducible testing conditions.

During preliminary tests [1] it has been found that the system latency is an important factor that influences the naturalness of the virtual environment. The system latency is the time passing between the subject's start of the phonation and the begin of the auralisation. To specify this influence we carried out a listening test. During this test, the participants were asked to judge the naturalness of their own speech signal, depending on an artificially added delay. As can be seen in figure 1, more than half of the tested persons considered their hearing impression as unnatural when the latency was above 23 ms.

The preliminary test results published at Euronoise 2003 [1] were carried out with commercially available software using standard reverberation algorithms that correspond to three virtual rooms with different reverberation time characteristics. Since the voice parameters that were analysed under these different room acoustical conditions did not differ significantly, it was decided to improve the setup and to include the influence of background noise in the measurements. This noise was expected to have significant influence on the speaker's conditions in reverberant rooms. The increase of the energy density with rising reverberation time in a diffuse sound field is specified in equation 1 using the approximation of Sabine [2].

$$w = \frac{4 \cdot P \cdot T}{0,163 \cdot c \cdot V} \tag{1}$$

Defining that P is the acoustic power of the background

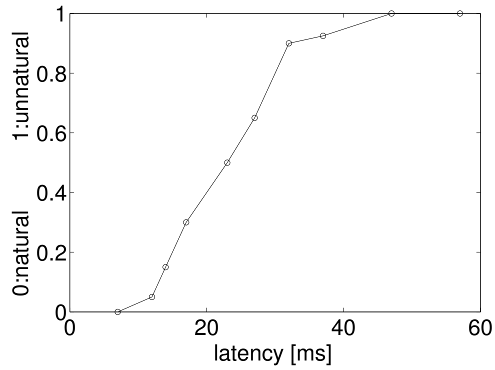

Figure 1: Naturalness vs. latency, average over all subjects

noise, c the speed of sound and V the volume of the performance room, we see that the energy density w is proportional to the reverberation time T.

Method

The setup can be divided into 3 parts, the recording, the processing and the auralisation. The speech signals were recorded in an anechoic chamber with about 10 cm space between mouth and microphone. After preamplification, the signal path is split. One branch is recorded directly for later analysis, and the other is mixed with background noise and used for simulation and auralisation. In figure 2 you can see a schematic flow-chart of the setup. The core of the simulation environment is a

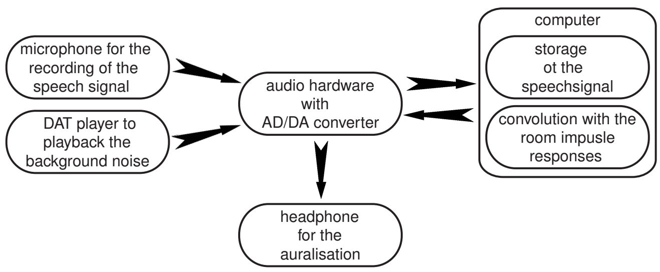

Figure 2: Signal flow of the set-up

very efficiently implemented convolution algorithm [4]. Depending on the performance of the deployed computer system, the software computes a real time convolution with room impulse responses spanning across a length between 0,7 (Pentium, 2 GHz) and 2,4 seconds (Athlon 2,2 GHz). The room impulse responses were measured by a special artificial head. Since this head can send an excitation signal through its mouth we can measure the transfer path from the mouth to the ears.

To accomplish a low system latency, as discussed above, we use a high performance audio hardware with ASIO drivers [6], which guarantees a fast data communication between the AD/DA converters and the software processing. With the deployed setup for the room simulation we achieved a total system latency of 11,3 ms. For auralisation we use an electrostatic open around-the-ear headphone (STAX SR). This provides a more natural hearing situation.

Tests

Up to now we ran two test series, the first one without any background noise and the second with shouting pupils as background sound. We recorded children attending the second class at elementary school during break time. In the first test the subjects were asked to read a text while the virtual environment of different rooms was simulated without background noise. For the second test series we added the described background noise to increase the acoustic effect of the virtual rooms.

In a second part of the test, we advised the subjects to deliver a spontaneous speech. This situation should be closer to the real speaking situation compared to the first part of the test. In both test series we simulated three different rooms that differ significantly in reverberation time. For the examination of stationary voice signals we also recorded a set of vowels.

Analysis and Results

During the analysis we followed two strategies, a subjective evaluation by a phoniatrician and an objective signal analysis by software. As objective speech quality parameters we examined pitch, sound pressure level and the duration of the text reading. In addition to these evaluations, we analysed the stationary vowel signals by help of the Göttingen Hoarseness Diagram and [5].

After interpreting the data of the first test we did not find any dependency of the voice quality on the room acoustics, neither with the objective evaluation nor with the subjective analysis. Also the evaluation of the data we recorded with the modified second setup did not result in significant changes of the objective evaluation of the voice quality parameters due to variations of the room acoustics (see figure 3). However, the subjective evaluation of the data we recorded with the modified second setup, gave interesting results. The phoniatrician heard variations in formant characteristic and attenuation of the voice timbre in those cases when the subject's voice characteristics exhibited minor pathologic properties (hyperfunctional voice). The direction of the variations and the room acoustical conditions were not correlated.

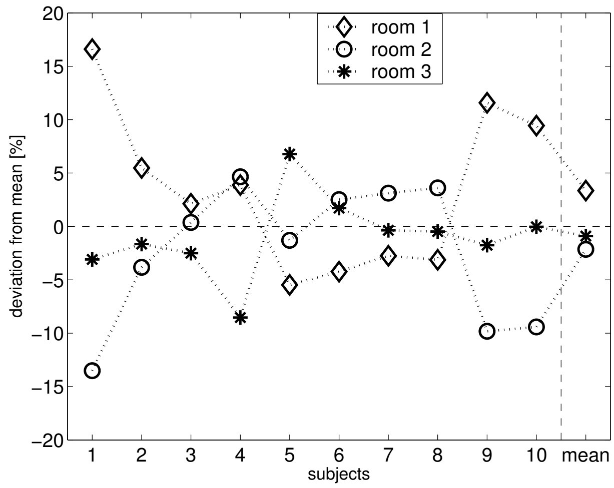

Figure 3: Pitch modulation during the readings
room 1 (T = 0,3 s); room 2 (T = 1,5 s); room 3 (T = 1,9 s)

Discussion

By way of conclusion we can say that the introduced virtual simulation provides a sufficiently natural acoustic environment, and as the convolution software can operate on any Windows PC, the environment can easily be set up. During the next steps of this project we have to rediscuss the accomplishment of the tests. One important topic will be the duration of the performed speeches, because the short length of the speech samples renders them useless to examine long time effects in the speech adaption. Apart from this, we also plan to improve the immersion of the subject into the classroom conditions, e.g. by extending the simulation with additional visual cues.

In future test series we want to narrow down the selection of the subjects to professional speakers and concentrate on teachers. Since in this occupational group many people suffer from voice disorders (about 30 %, described in [3]), we not only want to evaluate and diagnose the voice quality but also search for a training facility based on the room simulation environment.

References

[1] M. Kob, P. Heck, C. Neuschaefer-Rube (2003): Influence of room acoustics on voice quality. Proceedings of the 5^{th} European Conference on Noise Control – EURONOISE, CD-ROM file 429, 1-4.

[2] H. Kuttruff (2000): Room Acoustics. Taylor & Francis Group.

[3] I. Reinprecht (2002): Stimmtraining für Berufssprecher – eine Evaluation mittels Fragebogen. Diploma thesis at RWTH Aachen University.

[4] V. M. de Sousa da Graca (2000): PC-gestützte Echtzeitfaltung für raumakustische VR-Anwendungen. Diploma thesis at RWTH Aachen University.

[5] D. Michaelis, M. Fröhlich and H. W. Strube (1998): Selection and combination of acoustic features for the description of pathologic voices. J. Acoust. Soc. Am. 103(3), 1628-1640.

[6] Information at URL: `http://www.rme-audio.com`