



The technical Setup for Dialog Recordings in VMII and Problems caused by Mobile Phones

Karl Weilhammer
Christian Scheer

Ludwig Maximilians Universität München



Memo 128
December 1997

December 1997

Karl Weilhammer
Christian Scheer

Institut für Phonetik und Sprachliche Kommunikation
Ludwig Maximilians Universität München
Schellingstr. 3/II
80799 München

Tel.: (089) 2180 - 2806

e-mail: weilkar@phonetik.uni-muenchen.de

Gehört zum Antragsabschnitt: 6.1.4 Entwicklung von effizienten Aufnahme-
metechniken für beide Szenarien

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 G gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

Contents

1	Dialogues in VMII	2
2	The Technical Setup for Dialog Recordings in VMII	2
3	Problems with the recordings	3
3.1	Artifacts caused by the Laptop	3
3.2	Problems with interference caused by mobile phones	4
4	Methods to reduce mobile-phone interference	6
5	Summary	6

1 Dialogues in VMII

Monolingual dialogues for scenario 1 are recorded in Munich and Bonn. The task for the speakers is to fix a date for a one and a half days business trip to Hannover. Additionally they decide which means of transportation they want to use and choose a hotel to stay for the night. Finally they have a brief conversation about what they want to do in the evening after their meeting. Both dialog partners have a diary with some entries of fixed appointments, one has a timetable for flights and trains and the other one has a list of three hotels. Usually for each pair of speakers four of these dialogues are recorded. [techdok-59-97]

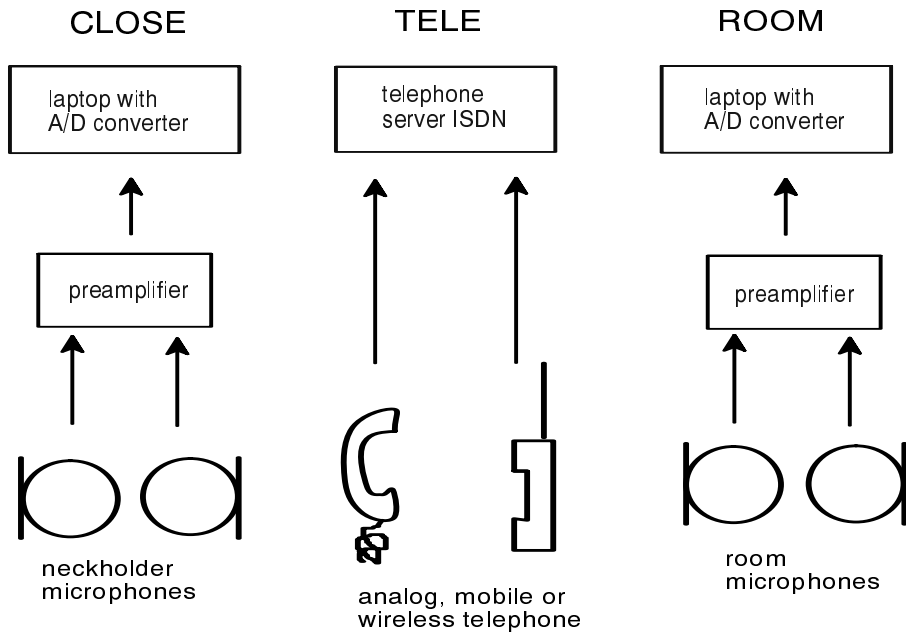


Figure 1: The three channels of recording for VMII dialogues.

2 The Technical Setup for Dialog Recordings in VMII

The recordings are made in a studio environment, where the dialog partners sit face to face at a table. Their voices are recorded in three ways (see also figure 1): Neck holder microphone (close), room microphone (room) and telephone (tele). Therefore three signal files are obtained for each speaker. The room and the close channel are recorded via preamplifiers on Toshiba laptops using an A/D converter with a sampling rate of 16 Bit at 16 kHz. The software was written in C for a Linux platform. The telephone channel is recorded by a PC

with a standard 2-channel ISDN card. The software was written in C++ based on CAPI. The sampling rate is 8 Bit at 8 kHz. For the telephone recordings standard analog telephones, mobile phones and an analog wireless phone are used.

The advantage of this technical setup is that it can work without external power supply. This provides the opportunity to record dialogues in different places, for instance in the environment of a business fair.

Recording directly on the hard disk of a laptop makes further processing of the data easier and faster.

Devices:

- 2 Toshiba laptops with Crystal CS 4243 A/D converter
- 2 Beyerdynamic MV100 preamplifiers
- 2 Beyer NEM 192 microphones (neck holder)
- 2 Beyer MCE 10 microphones (room)
- 1 AEG COMPACT EPS mobile telephone (distributed by e-plus)
- 1 Hagenuk st 900 sx wireless telephone
- 1 Actron B telephone (analog phone distributed by German Telecom)
- 1 PC with standard 2 channel ISDN card AVM/A1 (telephone server)

3 Problems with the recordings

3.1 Artifacts caused by the Laptop

Using the hard disk of a laptop as a storage medium for the acoustic signal has many advantages, but also causes some artifacts in the signal. The hard disk of a laptop does not spin all the time. If it is not needed for a while the motor that drives it is switched off. When the disk is needed again the motor is turned on again. This causes a decrease in the voltage supply of the A/D converter and therefore the signal voltage goes down as well. This phenomenon is easy to observe in the signal data by eye (Figure 2: in the middle of the 1st line). Since it is represented in the frequency range around 10 Hz it is hard to recognize in the acoustic signal by ear and should therefore not cause any problems to speech recognition systems.

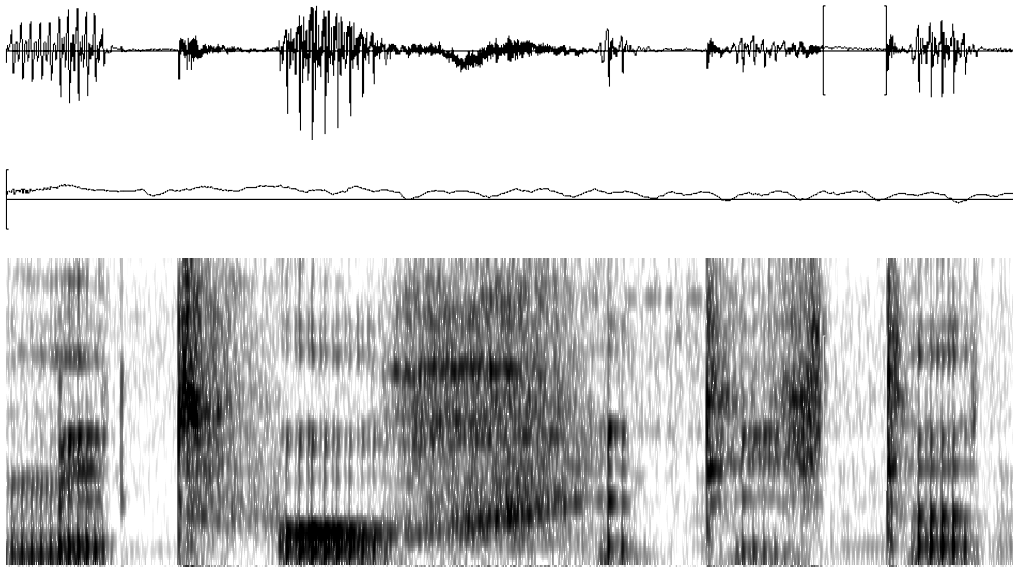


Figure 2: The acoustic signal vs time (first line and detail in brackets 2nd line) and sonogram (third line, 0 - 8 kHz) of a dialog sequence disturbed by hard disk artifacts.

Because of reasons similar to those described above, movements of the head of the disk cause very low noise on the signal. In the signal displayed in Figure 2 (1st line) one can compare the amplitude of the speech signal with the amplitude of a pause of the utterance where the artifact is acoustically observed (in brackets). The Detail (2nd line) shows this pause with a higher resolution. At the moment the recording software writes the data to disk every 5 seconds so that the low frequency noise occurs with a 5 second interval.

3.2 Problems with interference caused by mobile phones

In the VMII setup the dialog partners wear headsets and at least one of them uses a mobile phone. This means that the antenna of the latter is quite close to the wires that connect microphone, preamplifier and laptop, so that interference from the mobile phone's radiation cause artifacts on the recorded speech signal. Even the telephone line is affected. In the acoustic signal one either hears a permanent buzz or periodic clicks. This noise is sometimes so intense that it is even difficult to understand what was spoken in the dialog. We also tried recording a dialog with two people using mobile phones. As expected we found

interference in the close and room channel, but the telephone channel was clean.

If Verbmobil is to be used simultaneously with a mobile phone and a standard telephone, the speech recognition systems must be prepared to cope with this kind of noise.

As displayed in Figure 3 the interference are periodic artifacts that occur every 4 ms. It looks as if pieces of 0.5 ms duration have been cut out of the original Signal and then shifted downwards to result in such a discontinuous signal (1st line and detail in brackets in 2nd line). In the sonagram (3rd line) perturbed regions are characterized by a dense array of vertical lines. Unperturbed regions do not show these artifacts.

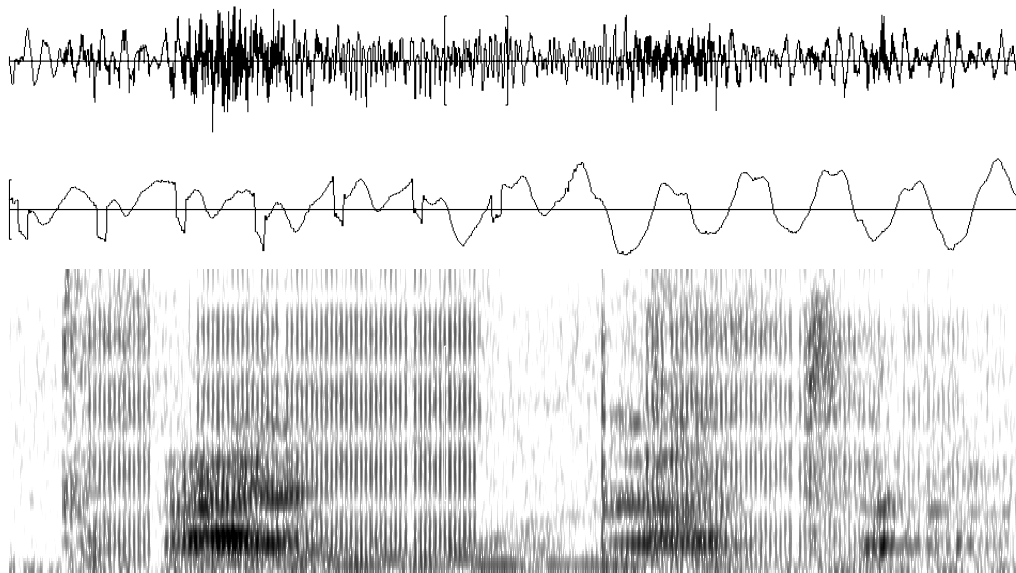


Figure 3: Acoustic signal vs time (1st line and detail in brackets 2nd line) and sonagram (third line, 0 - 8 kHz) of a dialog sequence with interference from a mobile phone. This signal is taken from a highly disrupted recording.

4 Methods to reduce mobile-phone interference

The frequency of mobile phones is in the GHz range. Electromagnetic waves are reflected by all kinds of metal, like metal doors, reinforced concrete ceilings or ventilation shafts. This makes it very difficult to shield electronic devices against interference, because it is very difficult to control reflections. Therefore we decided to get as far away with the mobile phone's antenna as possible and to remove big metal surfaces from the studio. We used a car antenna with a cable of 2,5 m and fixed it on a metal door to keep the biggest amount of the radiation outside the shielded studio. Nevertheless the internal antenna of the mobile phone still emits radiation to some extent. If the wires of the recording equipment are kept far away from the mobile phone or are shielded by the subject's body, interference can be effectively avoided.

As described in Section 1 we will deliver three sets of two dialog channels (close, room, tele). Only the close-channel will be without or at least almost without interference. Tele- and room-channel will have reduced interference because of the use of the car antenna.

5 Summary

The interference caused by radiation emitted by mobile phones are quite severe. We have found a method to record both clean and noisy training data, so that the designers of speech recognition systems will have a corpus that suits their demands.

But there still remains the problem that mobile phones cause interference in standard telephone lines. Even though the Verbmobil speech recognition systems will be able to cope with such artifacts, it is not agreeable for the users to have a noisy transmission of Verbmobil's translation.

If both users access Verbmobil by mobile phones no interference will occur. If both use standard phones problems will not occur either.

References

[techdok-59-97] Susanne Jekat, Christian Scheer; VMII Szenario I: Instruktionen; Universität Hamburg, LMU München.