Intelligibility and perceptual quality assessment of synthetic speech for RDS/TMC

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

15.02.1995

Rapport no. 1031

Intelligibility and perceptual quality assessment of synthetic speech for RDS/TMC. A proposal for evaluation

E.W.J. de Wit and L.L.M. Vogten

Intelligibility and perceptual quality assessment of synthetic speech for RDS/TMC. A proposal for evaluation.

E.W.J. de Wit and L.L.M. Vogten
Institute for Perception Research (IPO), Eindhoven, The Netherlands.

February, 1995

Contents

Abstract

New RDS/TMC systems are currently being developed for digital FM-broadcasting of traffic information, performing an automatic conversion of the received messages from ASCII-code into synthetic speech. This document provides and recommends an evaluation procedure for the synthetic speech component of TMC messages. Guidelines are given for the execution of the evaluation and a description is provided of the tools and resources needed for evaluation.

1. Introduction

The Traffic Message Channel (TMC), part of the Radio Data System (RDS), offers a new way for broadcasting updated traffic information. Combined with regular FM programs, TMC continuously transmits digitally coded traffic messages which are stored in receivers at home or in a car. In the RDS receiver the messages are decoded, displayed and automatically read aloud by means of synthetic speech. The car driver decides at which moment in time and how often the messages are spoken for the (sub)area of interest. There may also be a choice of language.

TMC messages are constructed of a carrier, one or more problem locations and an event. The event is the part of the message that specifies the nature of the traffic problem. Problem locations, frequently names of cities, exits etc., indicate the place where an event has occurred. The carrier of a message is composed by fixed parts and open slots to be filled with various problem locations and events.

Prestoring natural speech for each message would be far too impractical, mainly because of the large number of problem locations involved. Therefore synthetic speech will be generated in the RDS receiver by means of a "Phonemes to Speech" (PTS) conversion. Starting from an enriched phonetic description, appropriate short speech segments (diphones) are concatenated, prosodically modified by adjustment of duration and pitch, and then synthesized.

The phonetic transcriptions of carriers and events are stored in data bases within the RDS receiver, together with prosodic information for duration and intonation, copied from natural prosody. For the problem locations only the phonetic transcription is stored in an exchangeable chip card; the prosody is generated by means of rules.

Application of synthetic speech raises the question whether this type of speech is adequate to transfer TMC information to the user. Although intelligibility may be sufficient, synthetic speech is still clearly inferior to natural speech with respect to overall quality. In general, synthetic speech is more difficult to understand than natural speech (Pisoni, 1982, Eggen, 1992, Eggen and Nooteboom 1993).

The perceptual difference between natural and synthetic speech could be a major obstacle to the widespread use of synthetic speech in voice-output systems, particularly under conditions of high ambient noise or information overloading as may occur in a car environment.

The objective of this document is to provide and recommend an adequate evaluation procedure for the synthetic speech applied in TMC messages, to give guidelines for the execution of this evaluation and to describe the tools and resources needed for the evaluation.

The outline of this document is as follows. The next section gives some arguments and motivation for the choices that have been made for the current evaluation procedure. Section 3 gives further guidelines and specifications of the proposed procedure. After the references we present in, Appendix A1 to A6, all remaining details of the evaluation procedure.

2. Proposal

Ultimately it is the car driver who has to comprehend the synthetic speech. This requires subjective assessment by human listeners which means that we have to make allowance for the constraints of human observers. Pisoni et.al. (1985) mention a number of factors affecting an observer's performance.
The first factor that constrains performance concerns the complexity of the tasks that engage an observer during the perception of speech. Since a motorist has to pay close attention to the traffic when driving the car, this task could influence his performance on speech perception.
The second factor concerns the limitations in the human information processing system's ability to perceive, encode, store and retrieve information. For example, the amount of information that can be processed in and out of the short-term memory is severely limited by the listener's attentional state, past experience and the quality of the speech signal.
The third factor concerns the ability of human observers to quickly learn effective cognitive and perceptual strategies to improve performance in almost any task. Schwab et al. (1985) showed that perceptual training and exposure could change the strategies humans use in perceiving synthetic speech.
The fourth factor relates to the structure of the message set. Intelligibility scores are influenced by the linguistic properties of the test materials, such as familiarity and length of the stimuli. According to Owens (1961) words characterized by a greater familiarity, even to a slight degree, are more intelligible than words of lower familiarity. For the RDS/TMC application it means that names of large cities are likely to be more intelligible than names of relatively small villages. Also word length affects the intelligibility. Hirch et al. (1954) showed that the intelligibility of meaningful words is highly correlated with the number of syllables per word.
And the fifth factor refers to the acoustic-phonetic and prosodic structure of a synthetic utterance. Under adverse conditions, the intelligibility of synthetic speech may show serious degradation because of the lack of redundancy in the signal. Synthetic speech is very sensitive to distortion, e.g. masking noise (Pratt, 1986).

We propose an evaluation procedure consisting of two parts: first an opinion rating test to evaluate relevant aspects of the overall perceptual quality of synthesized TMC messages and second a more detailed articulation test to determine the intelligibility of synthesized problem locations. Both tests will be performed with natural and synthetic speech.

2.1. Opinion rating test

Opinion rating, which is a frequently used category-judgment method in psychometric measurements, is employed to assess the naturalness of speech quality. One advantage of the method is that different impairment factors can be assessed simultaneously. Also, the opinions of the listeners can be directly assessed and trained subjects are not required (Kitawaki, 1992). On the other hand, this method has some problems, since several factors affect opinion rating results: selection and order of presentation of the test conditions, subject selection and instructions to the subjects.

Opinion rating does not give an absolute measure since the scales used by the listeners are not calibrated and therefore it can be used only for rank-ordering the experimental conditions (Steeneken, 1992). For a more absolute evaluation, the use of a reference condition is required as an anchor. In the current test, natural speech without car noise serves as such a reference.

In the opinion rating test, a five-point scale for speech quality will be used, corresponding to: excellent, good, fair, poor and bad (CCITT, 1989). In accordance with the CCITT, the value 5 represents the best grade, value 1 the worst. The mean value of the ratings is called the Mean Opinion Score (MOS).

Some examples for natural speech of telephone bandwidth 3.2 kHz are: PCM coded with 64 kb/s it has a MOS score of 4.3; coded with 16 kb/s low-delay CELP a score of 4.0 and coded with 2.4 kb/s LPC the MOS drops to 2.5 (Jayant, 1992).

2.2. Articulation test

The articulation test is designed for the assessment of the intelligibility of speech signals. As opposed to the opinion rating test, no attention is paid to other speech quality aspects, such as naturalness.

An articulation test determines the percentage correctly heard words, syllables or phonemes. This percentage is the so-called articulation score.

If nonsense words are used with equal frequency of occurrence of all phonemes, the test results could also be presented as confusions between phonemes. Such a confusion matrix then provides useful diagnostic information for improving the performance of the speech synthesis system (Steeneken, 1992).

In the current test we restrict ourselves to names of existing locations. Therefore we will not go further into diagnostic aspects. These location names will be presented to the listeners, who have to write down what was heard.

2.3. Location names

Application-specific testing implies the use of task-specific test materials (Pols, 1989). In the current TMC application, the location names are the most critical parts of the message. They have the largest variety and their prosody is fully rule based.

As stated before, one of the factors determining intelligibility of speech is the length of the word to be recognized. In a pilot test (de Wit, 1995) with Dutch location names, it was found that the articulation score for names of locations with more than 10.000 inhabitants, increased significantly with the number of syllables in those names. But for names of the relatively small towns, with less than 10.000 inhabitants, no such effect occurred. Less familiar names of small towns seem to act more like nonsense words, which means that their intelligibility decreases with word length. So in the current test we make use of a balanced set of location names consisting of one, two and three or more syllables.

Another aspect of intelligibility is the listener's familiarity with locations; names of well-

known large cities are better recognized than those of unknown small villages. In the pilot test mentioned (de Wit, 1995) a similar effect was found. The mean articulation score for city names was about 20% to 40% higher for larger locations than for smaller ones.
In the current test we chose location names from two categories: cities with 10.000 or more inhabitants and cities with less than 10.000 inhabitants.
Of course familiarity is not only determined by the number of inhabitants. Topological knowledge of locations, which may be strongly listener dependent, is another factor that may affect the articulation score. Therefore at the end of the test each subject is asked to indicate whether the locations, used in the articulation test, have been heard of before.

2.4. Car noise

In a car environment speech may be partially masked by noise. As mentioned, synthetic speech is more vulnerable to noise than natural speech. So a relevant condition for the current test is the addition of noise.
Evaluation of speech in the real situation of a moving car has practical limitations and disadvantages. Therefore we propose to perform the test under more reproducible conditions, using recordings of car noise as made by Van de Voort (1987). Fig. 1 shows the mean energy spectrum of this type of noise.

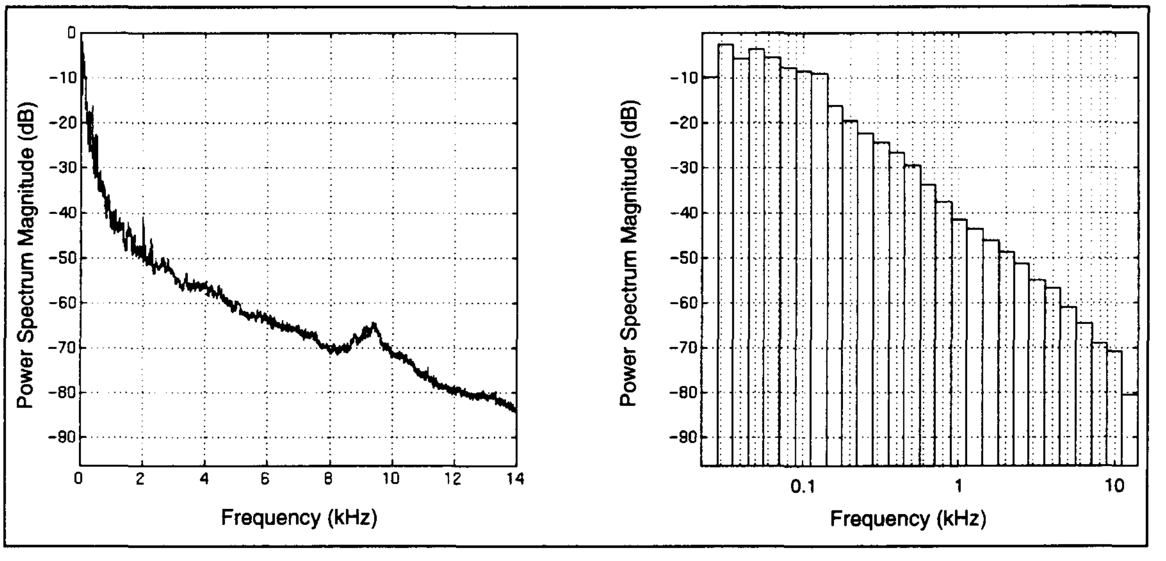

Fig. 1. Mean energy spectrum of the noise of a fairly small car, moving at 100 km/h (4th gear) on an asphalt highway, recorded by Van de Voort (1987). The spectra are determined from an interval of 30 seconds, with a window size of 8192 points. Left on a linear frequency scale, right on a logarithmic scale in third octave bands. Above 300 Hz the slope is about -6 dB/octave.

The speech and car noise will be presented to the listeners over headphones in order to avoid the acoustic influence of the listening room. Pilot tests (de Wit, 1995) have shown that the intelligibility of stimuli presented in an adverse noise condition over headphones or over loudspeakers was not significantly different.

2.5. Speech and noise levels

The speech level applied in the current evaluation is specified by a method adopted from

Steeneken and Houtgast (1986). This method is based on envelope detection and takes into account the silent intervals within a speech message, by introducing a threshold value before calculating the RMS-value. The result is called the RMSthr-value.
First the envelope of the speech signal is derived by a full-wave rectification followed by a 50 Hz low-pass filter. Then a level distribution histogram of the envelope function is calculated and from this histogram the RMSthr value is determined, omitting all envelope sample values below the threshold value. Figure 2 shows examples of envelope level histograms for natural speech, synthetic speech and car noise.
The RMS value of speech is of course dependent on pauses and "silent" segments within a speech message. As in the example in figure 2a is shown for natural speech, the level distribution histogram is composed of two distributions, one from the speech signal and one from the noise during "silent" periods, caused by microphone, amplifier, quantisation, etc. To separate these two components a threshold envelope value is introduced. Waveform segments with an envelope below this threshold are ignored, which reduces the contribution of the "silent" periods to the resulting RMS-value. Applying such a threshold to the speech envelope function rather than to the waveform means that above this threshold the contribution of the whole waveform is left intact, also the low-level parts in the "fine structure" near zero-crossings.
For synthetic speech the example in figure 2b shows that silent segments contributed only at zero level (-∞ dB). The level distribution histogram of the (quasi) stationary car noise is shown in figure 2c.
Before the calculation of the levels, a frequency weighting (A-filter) is applied to the speech signal, which is standard practice for acoustical measurements.

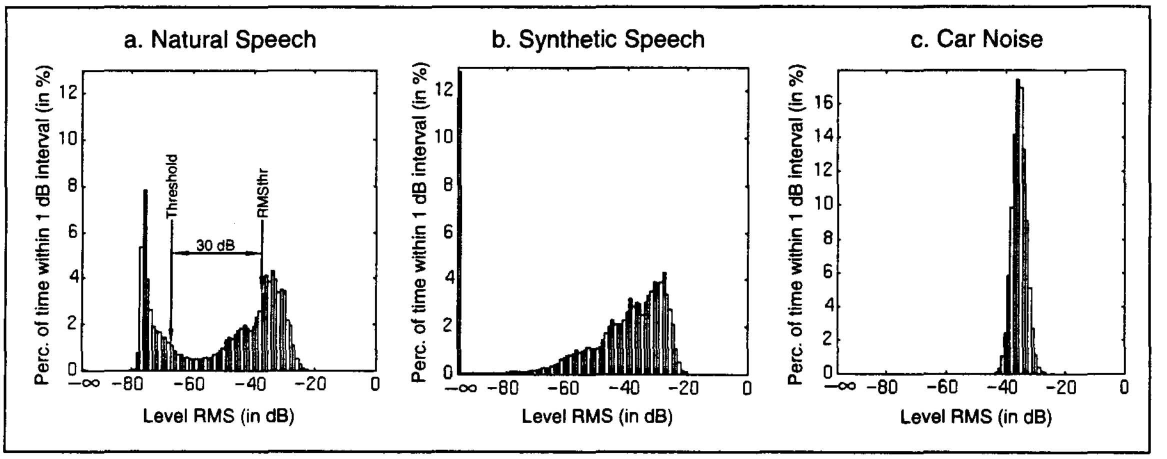

Fig. 2. Level distribution histogram of the envelope function obtained after full-wave rectifying and low pass filtering, according to Steeneken and Houtgast (1986). The speech material consisted of 60 seconds of spoken German traffic messages and the car noise is a 60 seconds recording from van de Voort (1987). The 0 dB level corresponds to the maximum 16 bit sample value.

To specify the actual speech and noise levels we use a Speech Level Meter developed at TNO Human Factors Research Institute in Soesterberg. This meter makes use of the algorithm as described above. Without applying a threshold the RMS level for the natural speech example in figure 2a is -39.0 dB and for synthetic -35.3 dB. With threshold the RMSthr for natural speech is -37.4 dB applying a 30 dB threshold at -67.4 dB in figure 2a.

For synthetic speech (fig. 2b) the RMSthr is -34.5 dB and for the car noise in (fig. 2c) we found an RMSthr of -35 dB. The value for the car noise is independent of the threshold applied, because there are no envelope values below the threshold value of -65 dB.

Levels of car noise and speech are set according to Van de Voort (1987). He found that the preferred listening level of speech in a car increases with background car noise level as shown in Fig. 3.

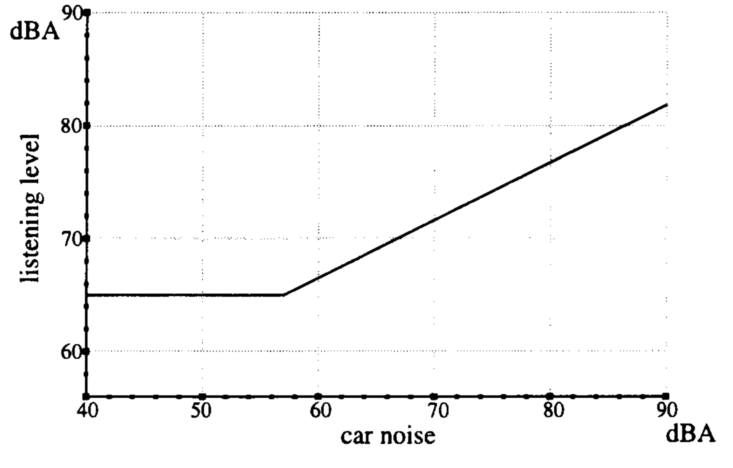

Fig. 3. Preferred speech level as a function of the background noise level according to van de Voort (1987). For noise below 57 dBA the listeners prefer a fixed speech level of 65 dBA. For higher noise levels the preferred speech level increases with 5 dB per 10 dB noise increment.

3. Evaluation of TMC synthesized speech

The first formal evaluation of RDS/TMC synthesized speech will be performed for German, but the test should also be applicable to other languages.

3.1. Speech versions, listening conditions and experimental set up.

Four speech versions are tested: natural and synthetic speech, both in the presence and absence of car noise. Natural and synthetic speech are both mono signals and have the same bandwidth of 5 kHz. The car noise is a stereo signal and has a bandwidth of 14 kHz. The car noise level chosen for the current test is 67 dBA. According to van de Voort (1987), the speech should be presented at 70 dBA, so 3 dB above car noise level.

Before addition of speech and car noise on DAT tape, the natural speech is scaled with 5.4 dB (factor 1.86) and the synthetic speech with 2.5 dB (factor 1.33), so that both have an RMSthr level of -32 dB or 3 dB above the RMSthr of -35 dB of the car noise.

We have chosen to present the stimuli over headphones to avoid the influence of room acoustics. The (stereo) car noise is presented dichotically (different signals to both ears) and the (mono) speech signal is presented diotically (same signal to both ears).

The stimuli are recorded on Digital Audio Tapes (DAT) and are played back from a DAT recorder (Sony TCD-D7), via a 6 channel headphone amplifier (Rane HC6) connected to 5 headphones (Sennheiser HD 530 II) for the listeners and one for monitoring purposes.
For the adjustment of the output level of the headphones, we make use of the Brüel & Kjær Precision Sound Level Meter Type 2203 in combination with the Brüel & Kjær Artificial Ear Type 4153. In the laboratory, DAT recorder output and amplifier are adjusted so that the car noise has a level of 67 dBA on each ear.
For calibration purposes a sinusoid of 1 kHz is recorded on each DAT-tape. At the location where the test takes place, DAT-recorder output and headphone amplifier are adjusted so that the amplitude of the sinusoid, measured on the headphones, is the same as in the laboratory situation.
For subjects with normal hearing a car noise level of 67 dBA corresponds to a sensation level of about 50 dB SL.

3.2. Listeners

The test is performed with 40 listeners, all native speakers of German and all having normal hearing. They are divided into four groups of 10 listeners, for methodological reasons further split up into two subgroups, A and B, of 5 listeners each.

3.3. Speech material

For the opinion rating the speech material consists of 16 passages, listed in appendix A2. Each passage is a concatenation of a few traffic messages with a total duration of about 25-30 seconds and is available in natural and in synthetic version. To each group of 5 listeners the passages are presented in the same order, but the conditions vary for each group.

For the articulation test, a set of 90 location names is selected, split up into two subsets of 45 names. This set is available in a natural and a synthetic version. The location names are embedded in short carrier phrases, functioning as a cue to the listener that the test word is about to be presented (Steeneken, 1992). Contents and meaning of the carrier phrases are not relevant and we use five different carrier phrases to keep the listener alert. A full list of the embedded locations names is given in appendix A4.
Appendix A4 also lists the names of all locations used, split up according to a (rather arbitrary) criterion of 10.000 inhabitants and according to the length of the names. The questionnaire to check the familiarity of all locations used is shown in Appendix A5.

3.4. Test procedure

Appendix A6 shows the schedule of stimulus presentation. The whole test consists of two sessions. Session 1 starts with a short introduction (5 min.) followed by the first opinion rating (6 min.), a short training session (2 min.) and the first articulation test (7 min.). Then, after a break, comes session 2 with the second opinion rating test (6 min.), the second articulation test (7 min.) and the questionnaire (3 min.) about familiarity.

3.4.1. Opinion rating

For the opinion rating a five-point scale is used in terms of "Ausgezeichnet", "Gut", "Mittelmäßig", "Dürftig", "Schlecht", with values of 5 down to 1. In Germany, however, people are accustomed to use the score '1' for the best school grade, which might cause confusions for the German listeners (Klaus et al., 1993). So, in the evaluation of German speech the order of the digits is reversed.

Subjects have to rate the passages on four aspects of the speech according to:
- overall perceptual quality
- acceptability for traffic messages
- intelligibility
- disturbing flaws in intonation and rhythm

The questions posed and the category labels of each aspect to be rated are listed in appendix A3. In each of the two sessions the subjects give their opinion about two of the four aspects. The subjects fill in the rating during or after the presentation of each passage; the response interval between two passages is 8 seconds.

3.4.2. Articulation test

Training

In order to acquaint the listeners with the response requirements, a short training session is run, consisting of 10 rather familiar locations embedded in a carrier. Each location has to be written down by the subjects. The training set of locations is not used in the main test and the speech version (natural or synthetic, with or without car noise) is the same as in the following main test. In appendix A4 the stimuli of this training session are listed.

Main test

The articulation test consists of two sessions. In each session 45 locations are tested, embedded in a short carrier. All listeners participate in both sessions and the subjects write down the location names they hear. The response interval after each stimulus is 8 seconds. To be able to check for possible learning effects, the 5 listeners of subgroup A receive list 1 in session 1 and list 2 in session 2. For subgroup B the presentation is the other way around, viz. list 2 in session 1 and list 1 in session 2. Each group of 10 subjects listens to only one of the four speech versions and each location is presented only once to each subject.

Scoring

The articulation score is expressed as the ratio of correctly heard locations and total number of locations presented in each session. Incorrect spelling, if representing correct phonemes, is counted as correct.

4. References

CCITT Recommendation P.80 (1989), Methods for Subjective Determination of Transmission Quality, *CCITT Blue Book, Vol. 5,* Geneva.

CCITT Supplement No. 2 (1989), Methods used for assessing telephony transmission performance, *CCITT Blue Book, Vol. 5,* Geneva.

Eggen, J.H. (1992), *On the Quality of Synthetic Speech. Evaluation and Improvements,* Doctoral Thesis TU Eindhoven.

Eggen, J.H. & Nooteboom, S.G. (1993), Speech quality and speaker characteristics. In: V.J. van Heuven & L.C.W. Pols (eds.): *Analysis and Synthesis of Speech, Strategic Research towards High-Quality Text-to-Speech Generation.* Mouton de Gruyter, Berlin, p. 279-288.

Hirch, I.J., Reynolds, E.G. & Joseph, M. (1954), Intelligibility of different speech materials, *Journal of the Acoustical Society of America, Vol. 26,* p. 530-538.

Jayant, N. (1992), High-quality coding of telephone speech and wideband audio, In: S. Furui and M.M. Sondhi (eds.): *Advances in speech signal processing.* M. Dekker Inc., New York, p. 92.

Kitawaki, N. (1992), Quality Assessment of Coded Speech, In: S. Furui and M.M. Sondhi, S. Furui and M.M. Sondhi (eds.): *Advances in speech signal processing.* M. Dekker Inc., New York, p. 357-385.

Klaus, H., Klix, H., Sotscheck, J. & Fellbaum, K. (1993), An Evaluation System for Ascertaining the Quality of Synthetic Speech Based on Subjective Category Rating Tests, *Proceedings of the 3rd European Conference on Speech Communication and Technology (Eurospeech '93),* Volume 3, Berlin, p. 1679-1682.

Owens, E. (1961), Intelligibility of words varying in familiarity, *Journal of Speech and Hearing Research, Vol. 4,* p. 113-129.

Pisoni, D.B. (1982), Perception of speech: The human listener as a cognitive interface, *Speech Technology 1,* p. 10-23.

Pisoni, D.B., Nusbaum, H.C. & Greene, B.G. (1985), Perception of synthetic speech generated by rule, *Proceedings of the IEEE, vol. 73,* p. 1665-1676.

Pols, L.C.W. (1989), Improving synthetic speech quality by systematic evaluation, *Speech input/output assessment and speech databases,* Workshop 20-23 Sept. Noordwijkerhout.

Pratt, R.L. (1986), On the intelligibility of synthetic speech, *Proc. Inst. of Acoustics, Vol 8, part 7,* p. 183-192.

Schwab, E.C, Nusbaum, H.C. & Pisoni, D.B. (1985), Some effect of training on the perception of synthetic speech, *Human Factors, 27(4),* p. 395-408.

Steeneken, H.J.M. & Houtgast, T. (1986), A comparison of some methods for measuring speech levels, *report IZF 1986-20,* Institute for perception TNO, Soesterberg, The Netherlands.

Steeneken, H.J.M. (1992), *On measuring and predicting speech intelligibility,* Doctoral thesis, University of Amsterdam.

Voort, van de A.T.A.M. (1987), Reproduction of Speech in Cars, *IPO-report no. 602.*

Wit de E.W.J. (1995), untitled, *IPO-report no. 1030.*

Appendices: Details of the test procedure

A1 Instruction sheet (in English and German)

Instructions for the listening test

In this experiment the quality of synthetic speech will be assessed. In the future this synthetic speech will be used when providing traffic information via a car radio. Besides synthetic speech you will also hear natural speech, both with and without car noise.

The experiment consists of two sessions, each with a duration of about 20 minutes.

First session

First you will get an opinion rating test. You will hear 8 spoken text passages each consisting three traffic messages. Each passage will be followed by a short pause in which you rate two aspects of the speech you have just heard. To that end you mark, on a five-point scale, the position which to your opinion applies best.

The second part consists of an intelligibility test. First you will get a training with 10 short traffic messages in which only one location name occurs. Please write down this location name on the corresponding form. After the training the main test follows with 45 short traffic messages. Please write clearly.

Second session

The second session again consists of an opinion rating test and an intelligibility test as in first session, but now without a preceding training session.
The session ends by filling in a list to indicate which location names you were familiar with before you took part in the test.

Thanks for your cooperation!

Anweisungen zum Hörtest

In diesem Test wird die Qualität von künstlicher Sprache für Verkehrsdurchsagen im Autoradio untersucht. Neben der künstlichen Sprache werden Sie auch natürliche Sprache beurteilen, beide sowohl mit als auch ohne Autogeräusch.

Der Test besteht aus zwei Teilen, die beide etwa 20 Minuten dauern.

Der erste Teil

Der erste Teil beginnt mit einem Bewertungstest. Dazu hören Sie eine Reihe von 8 Textpassagen. Jede Passage setzt sich aus drei Verkehrsdurchsagen zusammen. Nach jeder Passage folgt eine kurze Pause, in der Sie zwei Aspekte der soeben gehörten Sprachprobe beurteilen. Die Bewertung findet statt, indem Sie auf einem Fragebogen anhand einer 5-Punkte Skala angeben, welche Note Ihrer Meinung nach am besten zutrifft.

Als nächstes folgt ein Sprachverständlichkeitstest. Als Übung hören Sie zuerst 10 kurze Verkehrsdurchsagen, in denen jedesmal nur ein Ortsname vorkommt. Schreiben Sie diesen Ortsnamen auf den Fragebogen, und zwar bitte genau so, wie Sie ihn verstanden haben. Nach dieser Übung folgt dann der eigentliche Test mit 45 kurzen Verkehrsdurchsagen. Bitte schreiben Sie deutlich.

Der zweite Teil

Der zweite Teil besteht ebenfalls aus einem Bewertungstest und einem Verständlichkeitstest, aber nun ohne vorausgehende Übungsbeispiele. Nach diesem zweiten Teil geben Sie bitte auf dem Fragebogen an, welche der gehörten Ortsnamen Ihnen schon vor dem Test bekannt gewesen sind.

Herzlichen Dank für Ihre Mitarbeit.

A2 Speech material for the opinion rating

The following text passages are presented to the listeners in natural and in synthetic speech.

Passage 1:
A5, Karlruhe Richtung Frankfurt, Anschlußstelle Heidelberg: Orkanartige Stürme. Windgeschwindigkeit bis 150 km pro Stunde. A8, München Richtung Salzburg, zwischen Bad Aibling und Autobahndreiek Inntal: Nebel. B3, Celler Strasse, Hannover Richtung Celle, Anschlußstelle 5 Burgdorf: Unfall, 3 km Stau.

Passage 2:
A1, Nürnberg Richtung Passau, zwischen Nürnberg und Autobahnkreuz Altdorf: 4 km Stau. A3, Frankfurter Ring, Mainz Richtung Aschaffenburg, Anschlußstelle 51: Unfall. 6 km dichter Verkehr. A45, Offenbach Richtung Dortmund, zwischen Siegen und Anschlußstelle 18: nur 2 Fahrstreifen frei. Stau zu erwarten.

Passage 3:
A8, Stuttgart Richtung München, zwischen Günzburg und Burgau: 7 km Stau. A5, Karlruhe Richtung Frankfurt, zwischen Heidelberg und Autobahnkreuz Walldorf: Unfall. 5 km dichter Verkehr. B1, Berlin Richtung Magdeburg, Anschlußstelle Genthin: Orkanartige Stürme. Windgeschwindigkeit bis 150 km pro Stunde.

Passage 4:
A2, Dortmund Richtung Hannover, zwischen Rehren und Lauenau: 4 km Stau. A8, München Richtung Salzburg, zwischen Bad Aibling und Autobahndreiek 101: Unfall. 6 km dichter Verkehr. A65, Ludwigshafen Richtung Karlsruhe, zwischen Landau Zentrum und Insheim: bis zu 20 Minuten Zeitverluste zu erwarten.

Passage 5:
A4, Dresden Richtung Chemnitz, zwischen Wilsdruff und Autobahndreiek Nossen: Unfall. 6 km dichter Verkehr. A5, Karlruhe Richtung Frankfurt, zwischen Heidelberg und Walldorf: nur 2 Fahrstreifen frei. Stau zu erwarten. A63, Mainz Richtung Kaiserslautern, Anschlußstelle 12: stockender Verkehr.

Passage 6:
A3, Frankfurter Ring, Mainz Richtung Aschaffenburg, Anschlußstelle 51 Neu Isenburg: stockender Verkehr. A8, München Richtung Salzburg, zwischen Bernau und Anschlußstelle Felden: Unfall, 6 km dichter Verkehr. A72, Hof Richtung Chemnitz, Anschlußstelle Pirk, stockender Verkehr.

Passage 7:
A4, Dresden Richtung Chemnitz, zwischen Wilsdruff und Autobahndreiek 76: Unfall. 6 km dichter Verkehr. A6, Mannheim Richtung Saarbrücken, zwischen Neunkirchen und Homberg: Nebel. A13, Berlin Richtung Dresden, zwischen Lübben und Anschlußstelle Freiwalde: Unfall. 8 km dichter Verkehr.

Passage 8:
A1, Lübeck Richtung Hamburg, Autobahnkreuz 31 Hamburg Ost: Ausfahrt gesperrt. A8, Stuttgart Richtung München, zwischen Günzburg und Anschlußstelle Burgau: Unfall. 6 km dichter Verkehr. A72, Chemnitz Richtung Hof, zwischen Plauen Süd und Anschlußstelle 5: nur 2 Fahrstreifen frei. Stau zu erwarten.

Passage 9:
A4, Dresden Richtung Chemnitz, Autobahndreiek Nossen: Unfall. 6 km Stau. A8, Stuttgart Richtung München, Anschlußstelle Günzburg: stockender Verkehr. A65, Karlsruhe Richtung Ludwigshafen, zwischen Insheim und Anschlußstelle 16: bis zu 20 Minuten Zeitverluste zu erwarten.

Passage 10:
A8, München Richtung Salzburg, zwischen Bernau und Felden: Nebel. A65, Ludwigshafen Richtung Karlsruhe, Anschlußstelle 16 Landau Zentrum: Unfall. 6 km dichter Verkehr. A72, Hof Richtung Chemnitz, zwischen Pirk und Anschlußstelle 6: bis zu 20 Minuten Zeitverluste zu erwarten.

Passage 11:
A4, Dresden Richtung Chemnitz, Autobahndreiek 76 Nossen: Unfall. 6 km Stau. A13, Berlin Richtung Dresden, zwischen Freiwalde und Autobahndreiek Spreewald: 7 km Stau. A65, Karlsruhe Richtung Ludwigshafen, zwischen Insheim und Anschlußstelle Landau Zentrum: bis zu 20 Minuten Zeitverluste zu erwarten.

Passage 12:
A6, Mannheim Richtung Saarbrücken, zwischen Neunkirchen und Anschlußstelle Homberg: Nebel. A65, Ludwigshafen Richtung Karlsruhe, zwischen Landau Zentrum und Anschlußstelle 18: bis zu 20 Minuten Zeitverluste zu erwarten. B1, Berlin Richtung Magdeburg, Anschlußstelle Burg: Unfall. 3 km Stau.

Passage 13:
A1, Nürnberg Richtung Passau, zwischen Nürnberg und Autobahnkreuz 64: 2 km Stau. A8, München Richtung Salzburg, zwischen Bernau und Anschlußstelle 107: Unfall. 12 km dichter Verkehr. A72, Chemnitz Richtung Hof, zwischen Plauen Süd und Anschlußstelle Pirk: nur 2 Fahrstreifen frei. Stau zu erwarten.

Passage 14:
A3, Frankfurter Ring, Mainz Richtung Aschaffenburg, Anschlußstelle 51 Neu Isenburg: Unfall. 3 km Stau. A5, Karlruhe Richtung Frankfurt, zwischen Heidelberg und Autobahnkreuz 40: 7 Km Stau. A45, Offenbach Richtung Dortmund, zwischen Siegen und Anschlußstelle Olpe: Nebel.

Passage 15:
A1, Lübeck Richtung Hamburg, Autobahnkreuz Hamburg Ost: stockender Verkehr. A8, Stuttgart Richtung München, Anschlußstelle 67 Günzburg; stockender Verkehr. A13, Berlin Richtung Dresden, zwischen Lübben und Anschlußstelle 8: nur 2 Fahrstreifen frei. Stau zu erwarten.

Passage 16:
A4, Frankfurter Ring, Olpe Richtung Aachen, Autobahnkreuz 12: Ausfahrt gesperrt. A6, Mannheim Richtung Saarbrücken, zwischen Neunkirchen und Anschlußstelle 9: Nebel. A65, Ludwigshafen Richtung Karlsruhe, Anschlußstelle Landau Zentrum: Unfall. 6 km dichter Verkehr.

A3 Aspects to be evaluated in the opinion rating

Aspect 1: Overall perceptual quality

How do you rate the overall quality of the speech you have just heard?
5 = Excellent
4 = Good
3 = Fair
2 = Poor
1 = Bad

Wie beurteilen Sie die Gesamtqualität des Gehörten?
1 = Ausgezeichnet
2 = Gut
3 = Mittelmäßig
4 = Dürftig
5 = Schlecht

Aspect 2: Acceptability for traffic messages

Do you think the quality of the speech you just heard is suitable to provide traffic information in a car?
5 = Very suitable
4 = Suitable
3 = Doubtful
2 = Unsuitable
1 = Totally unsuitable

Finden Sie die Qualität der soeben gehörte Sprache geeignet für Verkehrsmeldungen im Auto?
1 = Sehr geeignet
2 = Geeignet
3 = Mittelmäßig
4 = Ungeeignet
5 = Völlig ungeeignet

Aspect 3: Intelligibility

Did you find certain words hard to understand?
5 = Never
4 = Rarely
3 = Occasionally
2 = Often
1 = All of the time

Hatten Sie Schwierigkeiten beim Verstehen bestimmter Wörter?
- 1 = Keine
- 2 = Selten
- 3 = Von Zeit zu Zeit
- 4 = Oft
- 5 = Ständig

Aspect 4: Intonation and rhythm

Did you notice any disturbing flaws in the intonation or rhythm of the speech messages?
- 5 = No flaws noticed
- 4 = Some flaws noticed, but not disturbing
- 3 = Slightly disturbing flaws noticed
- 2 = Disturbing flaws noticed
- 1 = Highly disturbing flaws noticed

Haben Sie störende Fehler bemerkt in der Betonung oder im Rhythmus der gehörten Meldungen?
- 1 = Keine Fehler bemerkt
- 2 = Einige Fehler bemerkt, die aber nicht stören
- 3 = Leicht störende Fehler bemerkt
- 4 = Störende Fehler bemerkt
- 5 = Sehr störende Fehler bemerkt

A4 List of location names used in the training and articulation test

Training session
The following carrier embedded locations are presented as training:

Im Bereich Frankfurt *Nebel*
Anschlußstelle Heidelberg *gesperrt*
Im Stadtgebiet Mannheim *Sportveranstaltung*
Richtung Hannover *Stau*
Im Regierungsbezirk Saarbrücken *Glatteis*
Im Bereich Hof *Nebel*
Anschlußstelle Aachen *gesperrt*
Im Stadtgebiet Mainz *Sportveranstaltung*
Richtung Berlin *Stau*
Im Regierungsbezirk Wuppertal *Glatteis*

Articulation test
Locations used for the articulation test, listed according to the number of inhabitants and according to the number of syllables of the location names.

< 10.000 inhabitants			≥ 10.000 inhabitants		
number of syllables			number of syllables		
1	2	>2	1	2	>2
Bann	Bogen	Adelsried	Bonn	Augsburg	Bielefeld
Cham	Dernbach	Appenweier	Fürth	Braunschweig	Bremerhaven
Daun	Eltmann	Bordesholm	Goch	Bremen	Düsseldorf
Haidt	Gernsheim	Emmelshausen	Hamm	Bayreuth	Erlangen
Holz	Kochel	Hermeskeil	Hürth	Darmstadt	Gelsenkirchen
Kenn	Landstuhl	Hollenstedt	Kiel	Duisburg	Ingolstadt
Kist	Leipheim	Iffeldorf	Köln	Essen	Leverkusen
Linz	Marktbreit	Ladbergen	Marl	Gladbeck	Oberhausen
Lorsch	Naunhof	Lobenstein	Moers	Jena	Oldenburg
Plech	Neumarkt	Medebach	Neuss	Koblenz	Osnabrück
Schlitz	Riegel	Merklingen	Soest	Krefeld	Paderborn
Strümp	Schnelldorf	Osterburken	Suhl	Leipzig	Regensburg
Tarp	Schwarmstedt	Sittensen	Trier	Potsdam	Rüsselsheim
Weil	Teupitz	Taufkirchen	Ulm	Rostock	Solingen
Zingst	Waldesch	Weibersbrunn	Worms	Wetzlar	Wiesbaden

The following carrier embedded locations are presented in the two sessions of the articulation test:

List 1

Im Bereich Trier *Nebel*
Anschlußstelle Weibersbrunn *gesperrt*
Im Stadtgebiet Leipzig *Sportveranstaltung*
Richtung Holz *Stau*
Im Regierungsbezirk Riegel *Glatteis*
Im Bereich Oberhausen *Nebel*
Anschlußstelle Wetzlar *gesperrt*
Im Stadtgebiet Neuss *Sportveranstaltung*
Richtung Emmelshausen *Stau*
Im Regierungsbezirk Dernbach *Glatteis*
Im Bereich Zingst *Nebel*
Anschlußstelle Rüsselsheim *gesperrt*
Im Stadtgebiet Schnelldorf *Sportveranstaltung*
Richtung Fürth *Stau*
Im Regierungsbezirk Essen *Glatteis*
Im Bereich Kist *Nebel*
Anschlußstelle Düsseldorf *gesperrt*
Im Stadtgebiet Hollenstedt *Sportveranstaltung*
Richtung Landstuhl *Stau*
Im Regierungsbezirk Iffeldorf *Glatteis*
Im Bereich Kiel *Nebel*
Anschlußstelle Rostock *gesperrt*
Im Stadtgebiet Osnabrück *Sportveranstaltung*
Richtung Haidt *Stau*
Im Regierungsbezirk Bayreuth *Glatteis*
Im Bereich Lorsch *Nebel*
Anschlußstelle Osterburken *gesperrt*
Im Stadtgebiet Waldesch *Sportveranstaltung*
Richtung Oldenburg *Stau*
Im Regierungsbezirk Goch *Glatteis*
Im Bereich Gladbeck *Nebel*
Anschlußstelle Köln *gesperrt*
Im Stadtgebiet Ladbergen *Sportveranstaltung*
Richtung Naunhof *Stau*
Im Regierungsbezirk Bielefeld *Glatteis*
Im Bereich Plech *Nebel*
Anschlußstelle Braunschweig *gesperrt*
Im Stadtgebiet Ulm *Sportveranstaltung*
Richtung Kochel *Stau*
Im Regierungsbezirk Appenweier *Glatteis*
Im Bereich Wiesbaden *Nebel*
Anschlußstelle Strümp *gesperrt*
Im Stadtgebiet Suhl *Sportveranstaltung*
Richtung Hermeskeil *Stau*
Im Regierungsbezirk Koblenz *Glatteis*

List 2

Im Bereich Soest *Nebel*
Anschlußstelle Bordesholm *gesperrt*
Im Stadtgebiet Darmstadt *Sportveranstaltung*
Richtung Tarp *Stau*
Im Regierungsbezirk Marktbreit *Glatteis*
Im Bereich Paderborn *Nebel*
Anschlußstelle Krefeld *gesperrt*
Im Stadtgebiet Sittensen *Sportveranstaltung*
Richtung Worms *Stau*
Im Regierungsbezirk Leipheim *Glatteis*
Im Bereich Weil *Nebel*
Anschlußstelle Erlangen *gesperrt*
Im Stadtgebiet Neumarkt *Sportveranstaltung*
Richtung Hamm *Stau*
Im Regierungsbezirk Bremen *Glatteis*
Im Bereich Kenn *Nebel*
Anschlußstelle Leverkusen *gesperrt*
Im Stadtgebiet Taufkirchen *Sportveranstaltung*
Richtung Schwarmstedt *Stau*
Im Regierungsbezirk Merklingen *Glatteis*
Im Bereich Hürth *Nebel*
Anschlußstelle Potsdam *gesperrt*
Im Stadtgebiet Regensburg *Sportveranstaltung*
Richtung Cham *Stau*
Im Regierungsbezirk Duisburg *Glatteis*
Im Bereich Bann *Nebel*
Anschlußstelle Lobenstein *gesperrt*
Im Stadtgebiet Gernsheim *Sportveranstaltung*
Richtung Bremerhaven *Stau*
Im Regierungsbezirk Moers *Glatteis*
Im Bereich Augsburg *Nebel*
Anschlußstelle Marl *gesperrt*
Im Stadtgebiet Adelsried *Sportveranstaltung*
Richtung Eltmann *Stau*
Im Regierungsbezirk Gelsenkirchen *Glatteis*
Im Bereich Schlitz *Nebel*
Anschlußstelle Bonn *gesperrt*
Im Stadtgebiet Jena *Sportveranstaltung*
Richtung Bogen *Stau*
Im Regierungsbezirk Medebach *Glatteis*
Im Bereich Solingen *Nebel*
Anschlußstelle Daun *gesperrt*
Im Stadtgebiet Ingolstadt *Sportveranstaltung*
Richtung Teupitz *Stau*
Im Regierungsbezirk Linz *Glatteis*

A5 Questionnaire about familiarity of the locations

Geben Sie bitte in der nachfolgenden Liste an, welche der Ortsnamen Ihnen schon vor Anfang dieser Tests bekannt gewesen sind und welche unbekannt.

bekannt	unbekannt	
❍	❍	Trier
❍	❍	Weibersbrunn
❍	❍	Leipzig
❍	❍	Holz
❍	❍	Riegel
❍	❍	Oberhausen
❍	❍	Wetzlar
❍	❍	Neuss
❍	❍	Emmelshausen
❍	❍	Dernbach
❍	❍	Zingst
❍	❍	Rüsselsheim
❍	❍	Schnelldorf
❍	❍	Fürth
❍	❍	Essen
❍	❍	Kist
❍	❍	Düsseldorf
❍	❍	Hollenstedt
❍	❍	Landstuhl
❍	❍	Iffeldorf
❍	❍	Kiel
❍	❍	Rostock
❍	❍	Osnabrück
❍	❍	Haidt
❍	❍	Bayreuth
❍	❍	Lorsch
❍	❍	Osterburken
❍	❍	Waldesch
❍	❍	Oldenburg
❍	❍	Goch

bekannt	unbekannt	
❍	❍	Gladbeck
❍	❍	Köln
❍	❍	Ladbergen
❍	❍	Naunhof
❍	❍	Bielefeld
❍	❍	Plech
❍	❍	Braunschweig
❍	❍	Ulm
❍	❍	Kochel
❍	❍	Appenweier
❍	❍	Wiesbaden
❍	❍	Strümp
❍	❍	Suhl
❍	❍	Hermeskeil
❍	❍	Koblenz
❍	❍	Soest
❍	❍	Bordesholm
❍	❍	Darmstadt
❍	❍	Tarp
❍	❍	Marktbreit
❍	❍	Paderborn
❍	❍	Krefeld
❍	❍	Sittensen
❍	❍	Worms
❍	❍	Leipheim
❍	❍	Weil
❍	❍	Erlangen
❍	❍	Neumarkt
❍	❍	Hamm
❍	❍	Bremen

bekannt	unbekannt	
❍	❍	Kenn
❍	❍	Leverkusen
❍	❍	Taufkirchen
❍	❍	Schwarmstedt
❍	❍	Merklingen
❍	❍	Hürth
❍	❍	Potsdam
❍	❍	Regensburg
❍	❍	Cham
❍	❍	Duisburg
❍	❍	Bann
❍	❍	Lobenstein
❍	❍	Gernsheim
❍	❍	Bremerhaven
❍	❍	Moers
❍	❍	Augsburg
❍	❍	Marl
❍	❍	Adelsried
❍	❍	Eltmann
❍	❍	Gelsenkirchen
❍	❍	Schlitz
❍	❍	Bonn
❍	❍	Jena
❍	❍	Bogen
❍	❍	Medebach
❍	❍	Solingen
❍	❍	Daun
❍	❍	Ingolstadt
❍	❍	Teupitz
❍	❍	Linz

A6 Test schedule

The tables below show which stimuli are presented to the different subgroups. Each subgroup consists of 5 listeners.

	group 1		group 2	
	subgroup A	subgroup B	subgroup A	subgroup B
Introduction	Reading of instruction sheet (see appendix A1)			
MOS-test (Session 1)	1 (N) 1-2 2 (N+n) 1-2 3 (S) 1-2 4 (S+n) 1-2 5 (N) 3-4 6 (N+n) 3-4 7 (S) 3-4 8 (S+n) 3-4	1 (N) 1-2 2 (S) 1-2 3 (N+n) 1-2 4 (S+n) 1-2 5 (N) 3-4 6 (S) 3-4 7 (N+n) 3-4 8 (S+n) 3-4	1 (S) 1-2 2 (S+n) 1-2 3 (N) 1-2 4 (N+n) 1-2 5 (S) 3-4 6 (S+n) 3-4 7 (N) 3-4 8 (N+n) 3-4	1 (S) 1-2 2 (N) 1-2 3 (S+n) 1-2 4 (N+n) 1-2 5 (S) 3-4 6 (N) 3-4 7 (S+n) 3-4 8 (N+n) 3-4
Training session	Training List (see appendix A4)			
Articulation test (Session 1)	list 1 (N)	list 2 (N)	list 1 (N+n)	list 2 (N+n)
Pause				
MOS-test (Session 2)	9 (N+n) 1-2 10 (N) 1-2 11 (S+n) 1-2 12 (S) 1-2 13 (N+n) 3-4 14 (N) 3-4 15 (S+n) 3-4 16 (S) 3-4	9 (N+n) 1-2 10 (S+n) 1-2 11 (N) 1-2 12 (S) 1-2 13 (N+n) 3-4 14 (S+n) 3-4 15 (N) 3-4 16 (S) 3-4	9 (S+n) 1-2 10 (S) 1-2 11 (N+n) 1-2 12 (N) 1-2 13 (S+n) 3-4 14 (S) 3-4 15 (N+n) 3-4 16 (N) 3-4	9 (S+n) 1-2 10 (N+n) 1-2 11 (S) 1-2 12 (N) 1-2 13 (S+n) 3-4 14 (N+n) 3-4 15 (S) 3-4 16 (N) 3-4
Articulation test (Session 2)	list 2 (N)	list 1 (N)	list 2 (N+n)	list 1 (N+n)
Questionnaire about familiarity	see appendix A5			

N=Natural speech S=Synthetic speech +n=with car noise

6 (S+n) 3-4 means, that text passage 6 will be presented to the listener in the 'synthetic speech with car noise' condition and questions about aspect 3 and 4 will be asked.

	group 3		group 4	
	subgroup A	subgroup B	subgroup A	subgroup B
Introduction	Reading of instruction sheet (appendix A1)			
MOS-test (Session 1)	1 (N+n) 1-2 2 (N) 1-2 3 (S+n) 1-2 4 (S) 1-2 5 (N+n) 3-4 6 (N) 3-4 7 (S+n) 3-4 8 (S) 3-4	1 (N+n) 1-2 2 (S+n) 1-2 3 (N) 1-2 4 (S) 1-2 5 (N+n) 3-4 6 (S+n) 3-4 7 (N) 3-4 8 (S) 3-4	1 (S+n) 1-2 2 (S) 1-2 3 (N+n) 1-2 4 (N) 1-2 5 (S+n) 3-4 6 (S) 3-4 7 (N+n) 3-4 8 (N) 3-4	1 (S+n) 1-2 2 (N+n) 1-2 3 (S) 1-2 4 (N) 1-2 5 (S+n) 3-4 6 (N+n) 3-4 7 (S) 3-4 8 (N) 3-4
Training session	Training List (see appendix A4)			
Articulation test (Session 1)	List 1 (S)	List 2 (S)	List 1 (S+n)	List 2 (S+n)
Pause				
MOS-test (Session2)	9 (N) 1-2 10 (N+n) 1-2 11 (S) 1-2 12 (S+n) 1-2 13 (N) 3-4 14 (N+n) 3-4 15 (S) 3-4 16 (S+n) 3-4	9 (N) 1-2 10 (S) 1-2 11 (N+n) 1-2 12 (S+n) 1-2 13 (N) 3-4 14 (S) 3-4 15 (N+n) 3-4 16 (S+n) 3-4	9 (S) 1-2 10 (S+n) 1-2 11 (N) 1-2 12 (N+n) 1-2 13 (S) 3-4 14 (S+n) 3-4 15 (N) 3-4 16 (N+n) 3-4	9 (S) 1-2 10 (N) 1-2 11 (S+n) 1-2 12 (N+n) 1-2 13 (S) 3-4 14 (N) 3-4 15 (S+n) 3-4 16 (N+n) 3-4
Articulation test (Session 2)	list 2 (S)	list 1 (S)	list 2 (S+n)	list 1 (S+n)
Questionnaire about familiarity	see appendix A5			