Joint SMG11/S4 **Tdoc SMG11/S4 ???**

Meeting: Kyoto, Japan 6-10 December 1999

Title : AMR-NS Selection Tests: Nortel Networks Test Results for Experiments 1, 4a, 4b, 5a, 5b, 6, 7

Source: **NORTEL NETWORKS**

Prepared on: **December 1, 1999**

Abstract: This report summarizes the results of listening tests carried out by Nortel Networks in support of the selection phase of the Adaptive Multi-Rate (AMR) Noise Suppresser (NS). A description of the recording process for source material prepared at Nortel is provided, along with details of the testing procedure as carried out at Nortel. The results of Experiment 1, 4a, 4b, 5a, 5b, 6, & 7 are presented in tabular form, comparing means and standard deviations obtained for candidates over the various test cases. Observations and comments about the experimental methodologies are provided.

Contact information:

Andrew Patrick
+1 613 765 7931
apatrick@nortelnetworks.com

Leigh Thorpe
+1 613 763 4382
thorpe@nortelnetworks.com

Summary of Deliverables for Contract

This report is a summary of the listening tests conducted by Nortel Networks under contract to ETSI as part of the SMG11/S4 AMR-NS selection program. This document describes the method of preparing source material supplied by Nortel for these tests, and the execution and results of listening tests completed with processed material provided by the host labs. Electronic copies of the completed analysis spreadsheet have been forwarded to the committee.

This report covers Experiments 1, 4a, 4b, 5a, 5b, 6, & 7 in the test plan. These experiments were conducted in accordance with the test plan (TDOC SMG11/S4 356/99 R, Issue 2.0), with the minor exceptions noted below. To complete these experiments, Nortel supplied source material to the host lab for processing through the various conditions. The processed material was returned to Nortel, and data were gathered from 24 listeners for each experiment (except Experiment 1, which involved 6 expert listeners). Groups of listeners heard the samples in the various orders as required by the test plan. Each listener participated in one experiment only.

Source Material Used at Nortel

Preparation of source material. The source material used in these experiments was prepared in-house to high quality standards. The recording method is described in detail below. Source speech from our laboratory has been used in many standardization and interworking tests conducted by TIA, ITU, and ETSI.

Recording equipment. The equipment making up the recording circuit included an AKG model #460B studio microphone equipped with microphone element #CK 61-ULS, a Symetrix model #SX202 pre-amplifier, and an Ariel model 656 ‘ProPort’ stereo audio-to-DSP port interface unit. The pre-amp was co-located with the microphone in the recording booth, and balanced output from the pre-amp fed the A/D converters outside the booth. A "Popper Stopper" windscreen was placed between the talker's mouth and the microphone to eliminate artifacts in the recordings. The microphone internal filter was set to 70 Hz to reduce any residual noise ‘rumble’ from air-conditioning equipment.

The recordings were made at a 16-kHz sampling rate using the anti-aliasing filter internal to the ProPort. The frequency response was flat between 75 Hz and 7.5 kHz; the upper cut-off was determined by the anti-aliasing filter.

The recording level was set to achieve the maximum signal-to-noise ratio with no amplitude clipping allowed for the duration of the speech file. The computer notified the operator if clipping was detected in the sample, in which case the talker was asked to do a second take.

The digital signal from the ProPort was sent to an Ariel DSP96002 DSP board resident in a Pentium computer. The speech data was copied to the Pentium hard drive.

Recording conditions. Talkers were all native speakers of English from eastern Ontario. Talkers were selected for the ability to read aloud, and were coached to read the sentences fluently, with natural (not dramatic) inflection.

The talker was seated in a double-walled sound attenuating chamber with an ambient noise level of 19 dBA. The interior of the chamber was treated with Sonex foam to control reverberation. Text to be spoken was presented to the talker using an LCD display to avoid problems with rustling paper. The microphone was placed about 8 inches from the talker's mouth, with the windscreen at about 4 inches.

The operator monitored each utterance, and requested a re-take in the event the sample was not acceptable. Acceptable samples were properly articulated, not subject to amplitude clipping or temporal clipping by the recording process, and free of artifacts such as mouth noise (e.g., whistles, pops) and background noise.

Editing and pre-processing of recorded samples. The recorded sentences were edited to standardize the leading and trailing silence around each sentence, and equalize the speech power across

the individual sentences. Following equalization, files were concatenated to create two-sentence samples. Because of the variation in sentence material, the samples varied in duration, but the samples were padded out to 8 sec using recordings of the ambient noise in the chamber through the recording circuit.

Pre-processing of the input samples (filtering, encoding, mixing of background noise into samples where required) was completed by the host lab.

Playback conditions: Variances from the test plan

Playback conditions were exactly as described in the test plan, with the following exceptions:

Presentation over headphones. Presentation of the samples was made over one channel of AKG stereo headphones, with the live transducer on the listener's telephone ear. The samples were filtered to reproduce the frequency characteristic of an ideal handset receiver at the listener's ear, calibrated at the ear reference point, before being played back over the headphone.

The use of a headphone reduces the variability of the heard frequency response compared to presentation over handset, because the headphone is relatively fixed on the listener's ear. When handsets are used, the listener is continually adjusting the ear seal from trial to trial; lower frequency components may leak out, and thus be absent from the heard signal. Headphones are more comfortable to wear over the course of the test session, and we can better ensure that the listener uses the same ear throughout the experiment. The main drawback with the use of headphones is that the opposite ear is also occluded, which is not the case with a handset. This becomes important for tests examining the effects of background noise in the listening environment.

A contribution describing the method of presentation of samples over headphone and especially the method of calibration of headphones for this use has been provided to ITU Study Group 12. (Nortel Networks (Canada), "Use of Headphones in Subjective Listening Studies of Telephone Transmission Quality", ITU-T SG12 Delayed Contribution D.94, Nov. 24 - Dec. 3, 1998, Geneva. Contact: L. Thorpe)

Reversing presentation order for some preliminary trials. The test plans for Experiments 4a, 4b, 5a, 5b, 6, and 7 did not always ensure that the preliminary (practice) trials included cases where the presentation was in the normal order (reference and then treated samples) and the reversed order (treated and then reference samples). Changes were made for each experiment so one half of the preliminary trials (randomly selected) had the samples presented in the reverse order.

Data Gathering and Manipulation

For all the experiments except Experiment 1, subjects made their responses by pressing one of seven response keys on a touch-sensitive computer screen. The responses were recorded by a laboratory computer for later processing and analysis. The raw data were coded numerically into a seven-point scale ranging from -3 ("much worse") to +3 ("much better"). These numerical assignments were reversed for those conditions where the processed sample was presented first. Finally, in Experiments 4a, 4b, 5a, and 5b, Conditions 4 and 5 from the test plan required special treatment since one third of the subjects received different background noise for these trials (producing conditions 4', 5', 4", and 5"). The raw data were loaded into Microsoft Excel (Version 97 SR-1) and SPSS (Version 9.0.1 for Windows) for the computation of means and standard deviations. Also, some additional analyses were performed to inspect rating bias as a function of the presentation order.

Experiment 1

Test Procedure

The procedure used for Experiment 1 followed the guidelines provided in the ETSI SMG11 AMR-NS/SQ test plan (Tdoc SMG11 288/99). A total of 96 conditions taken from Experiments 4, 5 and 2 were assessed by each of six expert listeners, who rated the six different Noise Suppression (NS) candidates with and without the different background noises and in the various processing conditions. The focus of Experiment 1 was on the initial convergence period and the artifacts that may occur during the convergence time of the candidate noise suppression techniques, during the leading silence, or at the time of speech onset.

Subjects. These data are based on a sample of six subjects. All six are highly experienced speech coding and DSP engineers. These individuals listen to and judge speech and audio signal quality on a day-to-day basis as part of their professional responsibilities.

Summary of Results

The results have been provided separately in an Excel document entitled AMR_NS_Nortel_Exp01_Data.xls, consisting of three tables. These include the means and standard deviations of ratings given on each question, for all talkers and for female talkers and male talkers separately (see Table 1.1 below). At a second stage, means for each question were computed for each of the six different Noise Suppressors (NS1, NS2, NS3, NS4, NS5, NS6) to provide an overall indication of performance with respect to each aspect judged (see Table 1.2 below). Finally, the comments made by each subject about the nature of any artifacts heard were presented in a third table for each of the 96 conditions (not included in this summary because of the large size).

Our preliminary analysis suggests an advantage for NS1, NS3 and NS6 regarding convergence time and artifacts during the leader silence. The absolute differences on Question 3 appear to suggest that NS1 and NS3 performed better on that rating, but the standard deviations for that data are larger, which suggests that the differences may not be statistically reliable.

Comments About the Test Method

Most of the subjects voiced a concern about the scale provided on Question 1. The categories used to indicate convergence were “not audible”, “audible, but fast”, or “audible and relatively slow”, leaving an ambiguity about the intended meaning for the choice of “not audible”. For example, if the noise suppressor was ineffective, there would be no audible convergence. This would be difficult to tell from a case where the NS converged very quickly on a sample with louder background noise. Inspection of cases where ratings and comments indicated that the rating of “1” apparently was meant to indicate the latter, we concluded that such cases made up only a small percentage of the entire set of scores. Statistically speaking, any discrepancies introduced by this ambiguity will not be detrimental to the interpretation of overall results.

Table 1.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 1 for Male and Female Talkers

					Question #1						Question #2						Question #3					
	Processed	Reference			Males Talkers		Females Talkers		All Talkers		Males Talkers		Females Talkers		All Talkers		Males Talkers		Females Talkers		All Talkers	
Condition	Codec/MNRU	Codec/MNRU	Noise	SNR value	Score	σ	Score	σ	Score	σ	Score	σ	Score	σ	Score	σ	Score	σ	Score	σ	Score	σ
1	AMR@5.9/NS1	AMR@5.9	Car	6 dB	1.67	1.15	1.00	0	1.33	0.82	1.67	1.15	2.00	1.00	1.83	0.98	2.00	0.00	1.33	0.58	1.67	0.52
2	AMR@5.9/NS2	AMR@5.9	Car	6 dB	2.00	0	2.33	1.15	2.33	0.52	1.50	1.00	2.00	1.00	1.67	1.03	1.50	0.58	1.67	0.58	1.50	0.55
3	AMR@5.9/NS3	AMR@5.9	Car	6 dB	1.25	0.5	1.67	1.15	1.50	0.84	1.00	0.00	1.33	0.58	1.00	0	1.50	0.58	2.00	0.00	1.67	0.52
4	AMR@5.9/NS4	AMR@5.9	Car	6 dB	2.25	0.96	2.00	0	2.33	0.52	1.75	0.96	1.33	0.58	1.67	0.82	1.75	0.50	2.33	1.15	2.00	0.89
5	AMR@5.9/NS5	AMR@5.9	Car	6 dB	2.50	1	3.00	0	3.00	0.00	1.25	0.50	1.67	1.15	1.50	0.84	2.00	0.00	1.67	1.15	1.83	0.75
6	AMR@5.9/NS6	AMR@5.9	Car	6 dB	1.25	0.5	1.33	0.58	1.33	0.52	1.25	0.50	1.33	0.58	1.33	0.52	2.00	0.82	2.00	0.00	2.00	0.63
7	AMR@5.9/NS1	AMR@5.9	Street	9 dB	1.25	0.5	1.67	1.15	1.50	0.84	1.00	0.00	1.33	0.58	1.17	0.41	1.50	0.58	1.00	0.00	1.17	0.41
8	AMR@5.9/NS2	AMR@5.9	Street	9 dB	1.75	0.96	1.75	0.5	2.00	0.63	1.50	0.58	1.50	1.00	1.67	0.82	1.75	0.50	1.75	0.96	1.67	0.82
9	AMR@5.9/NS3	AMR@5.9	Street	9 dB	1.75	0.96	1.00	0	1.50	0.84	1.25	0.50	1.25	0.50	1.33	0.52	1.50	0.58	2.00	0.82	1.67	0.82
10	AMR@5.9/NS4	AMR@5.9	Street	9 dB	2.50	1	2.00	1.15	2.67	0.82	1.50	0.58	1.50	1.00	1.67	0.82	1.75	0.50	1.25	0.50	1.33	0.52
11	AMR@5.9/NS5	AMR@5.9	Street	9 dB	2.00	1.15	2.00	1.15	2.33	1.03	1.25	0.50	1.50	1.00	1.50	0.84	1.25	0.50	1.25	0.50	1.00	0.00
12	AMR@5.9/NS6	AMR@5.9	Street	9 dB	1.75	0.96	1.25	0.5	1.33	0.52	1.50	0.58	1.25	0.50	1.50	0.55	1.75	0.96	1.75	0.50	1.83	0.75
13	AMR@5.9/NS1	AMR@5.9	Babble	9 dB	1.25	0.5	1.50	1	1.50	0.84	1.25	0.50	1.25	0.50	1.33	0.52	1.50	0.58	2.00	1.15	1.83	0.98
14	AMR@5.9/NS2	AMR@5.9	Babble	9 dB	1.50	0.58	1.75	0.96	1.83	0.75	1.75	0.50	1.25	0.50	1.67	0.52	1.25	0.50	1.50	1.00	1.17	0.41
15	AMR@5.9/NS3	AMR@5.9	Babble	9 dB	1.00	0	1.25	0.5	1.17	0.41	1.25	0.50	1.00	0.00	1.17	0.41	1.25	0.50	1.50	1.00	1.17	0.41
16	AMR@5.9/NS4	AMR@5.9	Babble	9 dB	1.50	1	1.75	0.96	1.83	0.98	1.50	0.58	1.25	0.50	1.33	0.52	1.25	0.50	1.75	0.50	1.67	0.52
17	AMR@5.9/NS5	AMR@5.9	Babble	9 dB	1.75	0.96	1.50	1	1.83	0.98	1.50	0.58	2.00	1.15	1.83	0.98	1.50	0.58	1.50	0.58	1.50	0.55
18	AMR@5.9/NS6	AMR@5.9	Babble	9 dB	1.00	0	1.25	0.5	1.17	0.41	1.50	1.00	1.50	0.58	1.67	0.82	2.25	0.50	1.50	0.58	2.00	0.63
19	AMR@5.9/NS1	AMR@5.9	Car	12 dB	1.50	0.58	1.25	0.5	1.50	0.55	1.00	0.00	1.00	0.00	1.00	0	1.50	0.58	1.50	0.58	1.50	0.55
20	AMR@5.9/NS2	AMR@5.9	Car	12 dB	2.00	0	1.75	0.5	1.83	0.41	1.25	0.50	1.00	0.00	1.17	0.41	1.50	0.58	1.50	0.58	1.50	0.55
21	AMR@5.9/NS3	AMR@5.9	Car	12 dB	2.00	0	1.50	0.58	1.67	0.52	1.00	0.00	1.00	0.00	1.00	0	1.50	0.58	1.50	0.58	1.50	0.55
22	AMR@5.9/NS4	AMR@5.9	Car	12 dB	2.25	0.5	2.00	0.82	2.33	0.52	1.50	0.58	1.50	0.58	1.67	0.52	1.25	0.50	1.50	0.58	1.17	0.41
23	AMR@5.9/NS5	AMR@5.9	Car	12 dB	2.75	0.5	2.25	0.96	2.83	0.41	1.25	0.50	1.25	0.50	1.33	0.52	2.25	0.50	1.50	0.58	2.00	0.63
24	AMR@5.9/NS6	AMR@5.9	Car	12 dB	2.00	0.82	1.25	0.5	1.33	0.52	1.00	0.00	1.00	0.00	1.00	0	2.75	0.50	1.50	0.58	2.00	0.89
25	AMR@5.9/NS1	AMR@5.9	Street	15 dB	2.00	0.82	1.00	0	1.17	0.41	1.00	0.00	1.00	0.00	1.00	0	1.75	0.96	1.50	0.58	1.33	0.52
26	AMR@5.9/NS2	AMR@5.9	Street	15 dB	1.75	0.5	1.75	0.96	1.83	0.75	1.50	1.00	1.25	0.50	1.50	0.84	1.75	0.96	1.25	0.50	1.17	0.41
27	AMR@5.9/NS3	AMR@5.9	Street	15 dB	1.00	0	1.25	0.5	1.17	0.41	1.00	0.00	1.00	0.00	1.00	0	1.75	0.96	1.50	0.58	1.67	0.82
28	AMR@5.9/NS4	AMR@5.9	Street	15 dB	1.75	0.96	1.75	0.96	2.00	0.89	1.25	0.50	1.50	0.58	1.50	0.55	1.25	0.50	1.50	0.58	1.33	0.52
29	AMR@5.9/NS5	AMR@5.9	Street	15 dB	1.75	0.96	2.00	1.15	2.17	0.98	1.25	0.50	1.00	0.00	1.17	0.41	1.25	0.50	1.50	0.58	1.33	0.52
30	AMR@5.9/NS6	AMR@5.9	Street	15 dB	1.00	0	1.25	0.5	1.17	0.41	1.50	1.00	1.00	0.00	1.33	0.82	1.25	0.50	1.75	0.96	1.50	0.84
31	AMR@5.9/NS1	AMR@5.9	Babble	15 dB	1.00	0	1.25	0.5	1.17	0.41	1.00	0.00	1.00	0.00	1.00	0	1.50	1.00	1.25	0.50	1.33	0.82

32	AMR@5.9/NS2	AMR@5.9	Babble	15 dB	1.25	0.5	1.25	0.5	1.33	0.52	1.50	0.58	1.50	0.58	1.67	0.52	1.75	0.96	1.50	0.58	1.33	0.52
33	AMR@5.9/NS3	AMR@5.9	Babble	15 dB	1.25	0.5	1.50	1	1.50	0.84	1.00	0.00	1.50	1.00	1.33	0.82	1.75	0.96	1.75	0.96	1.50	0.84
34	AMR@5.9/NS4	AMR@5.9	Babble	15 dB	1.50	1	1.50	1	1.67	1.03	1.25	0.50	1.25	0.50	1.33	0.52	1.75	0.50	1.25	0.50	1.50	0.55
35	AMR@5.9/NS5	AMR@5.9	Babble	15 dB	1.00	0	1.25	0.5	1.17	0.41	1.00	0.00	1.00	0.00	1.00	0	1.50	0.58	1.25	0.50	1.33	0.52
36	AMR@5.9/NS6	AMR@5.9	Babble	15 dB	1.25	0.5	1.00	0	1.17	0.41	1.50	0.58	1.25	0.50	1.50	0.55	1.50	0.58	1.50	0.58	1.33	0.52
37	AMR@12.2/NS1	AMR@12.2	Car	12 dB	1.25	0.5	1	0	1.17	0.41	1.25	0.5	1.75	0.96	1.67	0.82	1.5	0.58	2	0.82	1.67	0.82
38	AMR@12.2/NS2	AMR@12.2	Car	12 dB	1.75	0.5	2	0.82	2.17	0.41	1	0	1.25	0.5	1.17	0.41	2.25	0.5	2	0.82	2.17	0.75
39	AMR@12.2/NS3	AMR@12.2	Car	12 dB	1	0	1.5	0.58	1.33	0.52	1	0	1	0	1	0	1.75	0.5	1.5	0.58	1.5	0.55
40	AMR@12.2/NS4	AMR@12.2	Car	12 dB	2	1.15	2.25	0.96	2.5	0.84	1.5	0.58	2	0.82	2	0.63	2	0.82	1.5	0.58	1.67	0.82
41	AMR@12.2/NS5	AMR@12.2	Car	12 dB	2	1.15	2.75	0.5	2.83	0.41	1.25	0.5	1	0	1.17	0.41	1.75	0.5	1.5	1	1.5	0.84
42	AMR@12.2/NS6	AMR@12.2	Car	12 dB	1.25	0.5	2	0.82	1.5	0.55	1	0	1	0	1	0	2.5	0.58	1.75	0.5	2.33	0.52
43	AMR@12.2/NS1	AMR@12.2	Street	15 dB	1	0	1.75	0.96	1.17	0.41	1	0	1	0	1	0	1.5	0.58	1.25	0.5	1.33	0.52
44	AMR@12.2/NS2	AMR@12.2	Street	15 dB	1.75	0.96	1.5	0.58	1.83	0.75	1.25	0.5	1.25	0.5	1.33	0.52	1.25	0.5	1.5	1	1.33	0.82
45	AMR@12.2/NS3	AMR@12.2	Street	15 dB	1.25	0.5	1	0	1.17	0.41	1	0	1	0	1	0	1.25	0.5	1.25	0.5	1.17	0.41
46	AMR@12.2/NS4	AMR@12.2	Street	15 dB	2	1.15	2	0.82	2.33	0.82	1.5	1	1.5	1	1.67	1.03	1.5	0.58	1	0	1.17	0.41
47	AMR@12.2/NS5	AMR@12.2	Street	15 dB	2	1.15	2	1.15	2.33	1.03	1.5	1	1.25	0.5	1.5	0.84	1.25	0.5	1.25	0.5	1.17	0.41
48	AMR@12.2/NS6	AMR@12.2	Street	15 dB	1.5	0.58	1.25	0.5	1.5	0.55	1.5	1	1.25	0.5	1.5	0.84	2	0	1.5	0.58	1.83	0.41
49	AMR@12.2/NS1	AMR@12.2	Babble	15 dB	1.25	0.5	1	0	1.17	0.41	1	0	1	0	1	0	1.25	0.5	1	0	1	0
50	AMR@12.2/NS2	AMR@12.2	Babble	15 dB	1.25	0.5	1.5	1	1.5	0.84	1.25	0.5	1.25	0.5	1.33	0.52	1.25	0.5	1	0	1.17	0.41
51	AMR@12.2/NS3	AMR@12.2	Babble	15 dB	1	0	1.5	0.58	1.33	0.52	1	0	1	0	1	0	1.25	0.5	1	0	1	0
52	AMR@12.2/NS4	AMR@12.2	Babble	15 dB	1.25	0.5	1.5	1	1.5	0.84	1	0	1.25	0.5	1.17	0.41	1.25	0.5	1	0	1	0
53	AMR@12.2/NS5	AMR@12.2	Babble	15 dB	1.5	1	1.5	1	1.67	1.03	1	0	1.5	1	1.33	0.82	1.5	0.58	1.75	0.96	1.67	0.82
54	AMR@12.2/NS6	AMR@12.2	Babble	15 dB	1	0	1.25	0.5	1.17	0.41	1	0	1.25	0.5	1.17	0.41	1.25	0.5	1.25	0.5	1.33	0.52
55	AMR@12.2/NS1	AMR@12.2	Car	12 dB	1.75	0.96	1.25	0.5	1.67	0.82	1	0	1.5	0.58	1.33	0.52	1.25	0.5	1.25	0.5	1.33	0.52
56	AMR@12.2/NS2	AMR@12.2	Car	12 dB	2	0.82	2	0	2.17	0.41	1.25	0.5	1	0	1.17	0.41	1.25	0.5	1	0	1.17	0.41
57	AMR@12.2/NS3	AMR@12.2	Car	12 dB	1.75	0.5	1.25	0.5	1.5	0.55	1	0	1	0	1	0	1	0	1.25	0.5	1.17	0.41
58	AMR@12.2/NS4	AMR@12.2	Car	12 dB	2	0.82	1.75	0.96	2.17	0.75	1.25	0.5	1.25	0.5	1.33	0.52	1.25	0.5	1.5	0.58	1.33	0.52
59	AMR@12.2/NS5	AMR@12.2	Car	12 dB	2.75	0.5	2	1.15	2.83	0.41	1	0	1.25	0.5	1.17	0.41	1.75	0.96	1.5	0.58	1.67	0.82
60	AMR@12.2/NS6	AMR@12.2	Car	12 dB	1.5	0.58	1.25	0.5	1.33	0.52	1	0	1	0	1	0	2.75	0.5	1.5	0.58	2	0.89
61	AMR@12.2/NS1	AMR@12.2	Street	15 dB	1.75	0.5	1	0	1.33	0.52	1	0	1	0	1	0	2.25	0.96	1.5	0.58	1.67	0.82
62	AMR@12.2/NS2	AMR@12.2	Street	15 dB	1.5	0.58	1.75	0.96	1.67	0.82	1.75	0.96	1.25	0.5	1.67	0.82	1.5	1	1.25	0.5	1	0
63	AMR@12.2/NS3	AMR@12.2	Street	15 dB	1.25	0.5	1.75	0.5	1.5	0.55	1	0	1.25	0.5	1.17	0.41	1.75	0.96	1.5	0.58	1.33	0.52
64	AMR@12.2/NS4	AMR@12.2	Street	15 dB	1.5	0.58	2	1.15	1.83	0.98	1	0	1.75	0.5	1.5	0.55	1.5	1	1	0	1	0
65	AMR@12.2/NS5	AMR@12.2	Street	15 dB	2.5	0.58	2	1.15	2.5	0.84	1	0	1.25	0.5	1	0	2	1.15	1	0	1.33	0.82
66	AMR@12.2/NS6	AMR@12.2	Street	15 dB	1.25	0.5	1.75	0.96	1.33	0.52	1.25	0.5	1.25	0.5	1.17	0.41	1.5	0.58	1	0	1.33	0.52
67	AMR@12.2/NS1	AMR@12.2	Babble	15 dB	1	0	1.75	0.96	1.17	0.41	1.25	0.5	1.25	0.5	1.17	0.41	1.5	0.58	1	0	1.33	0.52
68	AMR@12.2/NS2	AMR@12.2	Babble	15 dB	1.5	0.58	1.25	0.5	1.5	0.55	1.5	0.58	1.5	0.58	1.5	0.55	1.5	0.58	1.25	0.5	1.33	0.52
69	AMR@12.2/NS3	AMR@12.2	Babble	15 dB	1.25	0.5	1.75	0.96	1.67	0.82	1.25	0.5	1.25	0.5	1.17	0.41	1.5	0.58	1.25	0.5	1.17	0.41

70	AMR@12.2/NS4	AMR@12.2	Babble	15 dB	1.25	0.5	1.75	0.96	1.67	0.82	1	0	1.25	0.5	1.17	0.41	1.25	0.5	1.25	0.5	1.17	0.41
71	AMR@12.2/NS5	AMR@12.2	Babble	15 dB	1	0	2	0.82	1.67	0.82	1	0	1	0	1	0	1.5	0.58	1.25	0.5	1.33	0.52
72	AMR@12.2/NS6	AMR@12.2	Babble	15 dB	1.25	0.5	1.25	0.5	1.33	0.52	1.25	0.5	1.25	0.5	1.33	0.52	1.75	0.5	1.25	0.5	1.5	0.55
73	AMR@12.2/NS1	AMR@12.2	Quiet	-	1.5	0.58	1.25	0.5	1.5	0.55	1.25	0.5	1.5	1	1.33	0.82	1.5	0.58	1.5	0.58	1.33	0.52
74	AMR@12.2/NS2	AMR@12.2	Quiet	-	1.5	1	1.75	0.5	1.67	0.82	2	0.82	2	0.82	2.33	0.52	1.25	0.5	1.25	0.5	1	0
75	AMR@12.2/NS3	AMR@12.2	Quiet	-	1	0	2.25	0.96	1.67	1.03	1.25	0.5	1.5	1	1.5	0.84	1.25	0.5	1.25	0.5	1	0
76	AMR@12.2/NS4	AMR@12.2	Quiet	-	1.5	1	1.5	1	1.67	1.03	1	0	2.25	0.96	1.83	0.98	1.25	0.5	1	0	1	0
77	AMR@12.2/NS5	AMR@12.2	Quiet	-	1	0	1	0	1	0	1	0	1.75	0.96	1.5	0.84	1.25	0.5	1	0	1	0
78	AMR@12.2/NS6	AMR@12.2	Quiet	-	1.5	1	1.5	1	1.67	1.03	2	0.82	1.75	0.96	2.17	0.75	1.25	0.5	1	0	1	0
79	AMR@5.9/NS1	AMR@5.9	Quiet	-	1	0	1.25	0.5	1.17	0.41	1	0	1	0	1	0	1.5	0.58	1.25	0.5	1.33	0.52
80	AMR@5.9/NS2	AMR@5.9	Quiet	-	1.5	1	1.75	0.96	1.83	0.98	1.75	0.96	1.75	0.96	2	0.89	1.25	0.5	1	0	1	0
81	AMR@5.9/NS3	AMR@5.9	Quiet	-	1.5	1	1	0	1.33	0.82	2	1.15	1	0	1.67	1.03	1.25	0.5	1	0	1	0
82	AMR@5.9/NS4	AMR@5.9	Quiet	-	1.5	1	1.5	1	1.67	1.03	2	0.82	1.25	0.5	1.83	0.75	1	0	1	0	1	0
83	AMR@5.9/NS5	AMR@5.9	Quiet	-	1	0	1.5	1	1.33	0.82	1.75	0.96	1	0	1.5	0.84	1	0	1	0	1	0
84	AMR@5.9/NS6	AMR@5.9	Quiet	-	1.5	1	1.5	1	1.67	1.03	1.75	0.96	1.25	0.5	1.67	0.82	1	0	1	0	1	0
85	AMR@12.2/NS1	AMR@12.2	Quiet	-	1.25	0.5	1	0	1.17	0.41	1	0	1	0	1	0	1.75	0.96	2	0.82	2.17	0.75
86	AMR@12.2/NS2	AMR@12.2	Quiet	-	1.25	0.5	1.5	0.58	1.5	0.55	1.25	0.5	1.25	0.5	1.33	0.52	1.25	0.5	1	0	1.17	0.41
87	AMR@12.2/NS3	AMR@12.2	Quiet	-	1	0	1	0	1	0	1	0	1.5	1	1.33	0.82	1	0	1	0	1	0
88	AMR@12.2/NS4	AMR@12.2	Quiet	-	1.5	1	1.25	0.5	1.5	0.84	1	0	1	0	1	0	1	0	1.25	0.5	1.17	0.41
89	AMR@12.2/NS5	AMR@12.2	Quiet	-	1.25	0.5	1	0	1.17	0.41	1.5	1	1	0	1.33	0.82	1	0	1	0	1	0
90	AMR@12.2/NS6	AMR@12.2	Quiet	-	1.25	0.5	1.5	1	1.5	0.84	1.75	0.5	1.25	0.5	1.67	0.52	1	0	1	0	1	0
91	AMR@5.9/NS1	AMR@5.9	Quiet	-	1.5	0.58	1.25	0.5	1.5	0.55	1	0	1	0	1	0	2	0.82	1.5	1	2	0.89
92	AMR@5.9/NS2	AMR@5.9	Quiet	-	2	0.82	1.25	0.5	1.67	0.82	1.25	0.5	1	0	1.17	0.41	1.5	0.58	1.25	0.5	1.33	0.52
93	AMR@5.9/NS3	AMR@5.9	Quiet	-	1.75	0.96	1	0	1.33	0.82	1.5	0.58	1.5	1	1.5	0.84	1.5	0.58	1.25	0.5	1.17	0.41
94	AMR@5.9/NS4	AMR@5.9	Quiet	-	1.5	0.58	1.5	1	1.5	0.84	1.25	0.5	1.75	0.96	1.67	0.82	1.5	0.58	1.75	0.5	1.5	0.55
95	AMR@5.9/NS5	AMR@5.9	Quiet	-	1.25	0.5	1.5	1	1.33	0.82	1	0	1.75	0.96	1.5	0.84	1.25	0.5	1.25	0.5	1	0
96	AMR@5.9/NS6	AMR@5.9	Quiet	-	1.75	0.96	1.5	1	1.67	1.03	1.25	0.5	1.5	0.58	1.5	0.55	1.25	0.5	1	0	1	0

Table 1.2: Overall Mean Opinion Scores (and Standard Deviations) For Each Candidate in Experiment 1

Codec	Mean (Q1)	Std. (Q1)	Mean (Q2)	Std. (Q2)	Mean (Q3)	Std. (Q3)
NS1	1.32	0.1863074	1.20	0.3667529	1.42	0.2359503
NS2	1.82	0.1859576	1.56	0.206786	1.32	0.2782214
NS3	1.43	0.2064296	1.17	0.3726708	1.32	0.2868242
NS4	2.01	0.1671562	1.55	0.2007582	1.31	0.2957176
NS5	2.11	0.3741606	1.32	0.3557572	1.40	0.3114659
NS6	1.36	0.2040254	1.38	0.3109901	1.64	0.2768276

Experiment 4a

Observations on the Test Procedure

Subjects in Experiments 4a, 4b, 5a, and 5b commented during the experiment that the frequent repetition of sample sentences was somewhat annoying, and might lead to difficulty separating the trials.

In addition, subjects often reported that they found the experiment too long, even with break periods during the session.

Replacing subjects due to randomization errors in the test plan. There was a problem in the random sequences included in the test plan that caused one test condition to be omitted and one to be tested twice as often as was called for. The affected data was discarded and replaced by data gathered with the corrected orders received from COMSAT.

Replacing subjects due to extreme data. Initial analysis of the results revealed fairly large differences between the ratings provided during normal-order trials (listen to reference and then processed and rate the processed) and reversed-order trials (listen to processed and then reference and rate the reference). For example, subjects' mean ratings for Conditions 6-32 in the test plan differed by 0.68 for normal and reversed-order presentations. At the request of group secretary, Paolo Usai, the results from each subject were examined in detail. This examination involved looking for evidence that subjects may have misunderstood the instructions and not rated the second stimulus in the reversed-order trials. Individual results were also examined to look for "outliers"; data that was clearly outside of the pattern created by the other subjects. Three subjects were identified as having extreme differences between their normal-order and reversed-order trials, and these subjects were replaced. Following these replacements, the mean difference for normal and reversed-order trials across the 24 subjects was 0.59.

Results

Table 4a.1 provides means and standard deviations of the final sample for Experiment 4a. Table 4a.2 preserves the information about presentation order. The results of the experiment have been reported in the soft copy of the spreadsheet provided to John Tardelli for the global analysis.

Table 4a.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 4a for Male and Female Talkers

						Male Talkers		Female Talkers		All Talkers	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	-0.052	0.686	0.042	0.695	-0.005	0.690
2	2	Street	AMR@x	-	AMR@x	0.010	0.718	0.146	0.665	0.078	0.693
3	3	Babble	AMR@x	-	AMR@x	0.052	0.701	0.010	0.788	0.031	0.744
4	4	Car	MNRU-16	-	MNRU-12	-0.563	0.759	-0.594	0.875	-0.578	0.813
5	5	Car	Direct	-	MNRU-12	-2.125	0.907	-1.844	0.987	-1.984	0.951
6	4'	Street	MNRU-16	-	MNRU-12	-0.969	1.092	-0.750	1.320	-0.859	1.207
7	5'	Street	Direct	-	MNRU-12	-1.813	1.203	-2.250	0.803	-2.031	1.038
8	4"	Babble	MNRU-16	-	MNRU-12	-1.000	0.984	-0.875	1.157	-0.938	1.067
9	5"	Babble	Direct	-	MNRU-12	-2.438	0.669	-2.500	0.622	-2.469	0.642
10	6	Car	AMR@x	4	AMR@x	0.604	0.827	0.427	0.707	0.516	0.772
11	7	Car	AMR@x	7	AMR@x	0.740	0.874	0.740	0.837	0.740	0.853
12	8	Car	AMR@x	10	AMR@x	1.083	1.053	1.115	0.893	1.099	0.974
13	9	Street	AMR@x	4	AMR@x	0.333	0.790	0.302	0.872	0.318	0.830
14	10	Street	AMR@x	7	AMR@x	0.688	0.987	0.667	0.879	0.677	0.932
15	11	Street	AMR@x	10	AMR@x	1.063	0.805	0.990	1.119	1.026	0.973
16	12	Babble	AMR@x	4	AMR@x	0.406	0.913	0.344	0.892	0.375	0.901
17	13	Babble	AMR@x	7	AMR@x	0.615	0.956	0.615	1.030	0.615	0.991
18	14	Babble	AMR@x	10	AMR@x	0.906	1.057	0.938	1.064	0.922	1.058
19	15	Car	AMR@x	-	AMR/NS1@x	0.521	1.095	0.625	1.069	0.573	1.080
20	16	Car	AMR@x	-	AMR/NS2@x	0.760	1.140	0.583	1.033	0.672	1.089
21	17	Car	AMR@x	-	AMR/NS3@x	0.531	1.046	0.479	1.026	0.505	1.033
22	18	Car	AMR@x	-	AMR/NS4@x	0.521	1.056	0.563	1.230	0.542	1.143
23	19	Car	AMR@x	-	AMR/NS5@x	0.708	1.095	0.500	0.962	0.604	1.033
24	20	Car	AMR@x	-	AMR/NS6@x	0.750	1.170	0.500	1.298	0.625	1.239
25	21	Street	AMR@x	-	AMR/NS1@x	0.292	1.045	0.490	0.951	0.391	1.002
26	22	Street	AMR@x	-	AMR/NS2@x	0.510	1.046	0.677	1.031	0.594	1.039
27	23	Street	AMR@x	-	AMR/NS3@x	0.500	1.056	0.406	1.052	0.453	1.052
28	24	Street	AMR@x	-	AMR/NS4@x	0.458	1.151	0.510	1.046	0.484	1.097
29	25	Street	AMR@x	-	AMR/NS5@x	0.177	0.929	0.365	0.942	0.271	0.938
30	26	Street	AMR@x	-	AMR/NS6@x	0.729	1.244	0.573	1.464	0.651	1.357
31	27	Babble	AMR@x	-	AMR/NS1@x	0.458	0.928	0.292	0.917	0.375	0.924
32	28	Babble	AMR@x	-	AMR/NS2@x	0.354	1.036	0.167	0.959	0.260	1.000
33	29	Babble	AMR@x	-	AMR/NS3@x	0.198	0.958	0.292	0.832	0.245	0.896
34	30	Babble	AMR@x	-	AMR/NS4@x	0.417	1.092	0.427	1.054	0.422	1.070
35	31	Babble	AMR@x	-	AMR/NS5@x	0.167	0.842	0.156	0.701	0.161	0.773
36	32	Babble	AMR@x	-	AMR/NS6@x	0.542	1.151	0.510	1.016	0.526	1.083

Table 4a.2: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 4a for Normal and Reversed Presentation Orders

						Normal Order		Reversed Order		Both Orders	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	-0.156	0.730	0.146	0.615	-0.005	0.690
2	2	Street	AMR@x	-	AMR@x	0.042	0.679	0.115	0.709	0.078	0.693
3	3	Babble	AMR@x	-	AMR@x	0.198	0.690	-0.135	0.763	0.031	0.744
4	4	Car	MNRU-16	-	MNRU-12	-0.875	0.871	-0.281	0.634	-0.578	0.813
5	5	Car	Direct	-	MNRU-12	-2.344	0.745	-1.625	1.008	-1.984	0.951
6	4'	Street	MNRU-16	-	MNRU-12	-1.625	1.070	-0.094	0.777	-0.859	1.207
7	5'	Street	Direct	-	MNRU-12	-2.219	0.975	-1.844	1.081	-2.031	1.038
8	4''	Babble	MNRU-16	-	MNRU-12	-1.531	0.915	-0.344	0.865	-0.938	1.067
9	5''	Babble	Direct	-	MNRU-12	-2.531	0.621	-2.406	0.665	-2.469	0.642
10	6	Car	AMR@x	4	AMR@x	0.625	0.771	0.406	0.762	0.516	0.772
11	7	Car	AMR@x	7	AMR@x	0.958	0.820	0.521	0.833	0.740	0.853
12	8	Car	AMR@x	10	AMR@x	1.490	1.005	0.708	0.767	1.099	0.974
13	9	Street	AMR@x	4	AMR@x	0.521	0.870	0.115	0.738	0.318	0.830
14	10	Street	AMR@x	7	AMR@x	0.896	1.010	0.458	0.794	0.677	0.932
15	11	Street	AMR@x	10	AMR@x	1.375	0.909	0.677	0.912	1.026	0.973
16	12	Babble	AMR@x	4	AMR@x	0.802	0.890	-0.052	0.686	0.375	0.901
17	13	Babble	AMR@x	7	AMR@x	1.167	0.914	0.063	0.723	0.615	0.991
18	14	Babble	AMR@x	10	AMR@x	1.583	0.914	0.260	0.729	0.922	1.058
19	15	Car	AMR@x	-	AMR/NS1@x	0.583	1.167	0.563	0.993	0.573	1.080
20	16	Car	AMR@x	-	AMR/NS2@x	0.802	1.228	0.542	0.917	0.672	1.089
21	17	Car	AMR@x	-	AMR/NS3@x	0.625	1.069	0.385	0.988	0.505	1.033
22	18	Car	AMR@x	-	AMR/NS4@x	0.563	1.263	0.521	1.015	0.542	1.143
23	19	Car	AMR@x	-	AMR/NS5@x	0.875	1.145	0.333	0.829	0.604	1.033
24	20	Car	AMR@x	-	AMR/NS6@x	0.823	1.369	0.427	1.064	0.625	1.239
25	21	Street	AMR@x	-	AMR/NS1@x	0.583	1.063	0.198	0.902	0.391	1.002
26	22	Street	AMR@x	-	AMR/NS2@x	0.885	1.104	0.302	0.884	0.594	1.039
27	23	Street	AMR@x	-	AMR/NS3@x	0.625	1.190	0.281	0.867	0.453	1.052
28	24	Street	AMR@x	-	AMR/NS4@x	0.792	1.169	0.177	0.929	0.484	1.097
29	25	Street	AMR@x	-	AMR/NS5@x	0.458	1.045	0.083	0.777	0.271	0.938
30	26	Street	AMR@x	-	AMR/NS6@x	1.073	1.453	0.229	1.110	0.651	1.357
31	27	Babble	AMR@x	-	AMR/NS1@x	0.844	0.898	-0.094	0.682	0.375	0.924
32	28	Babble	AMR@x	-	AMR/NS2@x	0.708	1.025	-0.188	0.744	0.260	1.000
33	29	Babble	AMR@x	-	AMR/NS3@x	0.656	0.904	-0.167	0.675	0.245	0.896
34	30	Babble	AMR@x	-	AMR/NS4@x	0.844	1.146	0.000	0.795	0.422	1.070
35	31	Babble	AMR@x	-	AMR/NS5@x	0.438	0.818	-0.115	0.613	0.161	0.773
36	32	Babble	AMR@x	-	AMR/NS6@x	0.979	1.205	0.073	0.700	0.526	1.083

Experiment 4b

Observations on the Test Procedure

Replacing subjects due to extreme data. Following the procedure adopted for Experiment 4a, the results from each subject were examined in detail. Two subjects were identified as having extreme differences between their normal-order and reversed-order trials, and these subjects were replaced. Following this replacement, the mean difference between normal and reversed-order trials for Conditions 6-32 in the test plan was 0.39.

Results

Table 4b.1 summarizes the final results of Experiment 4b in terms of means and standard deviations for each condition, preserving the information about talker gender. Table 4b.2 preserves the information about presentation order. Also, the results of the experiment have been reported in the soft copy of the spreadsheet provided by John Tardelli.

Table 4b.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 4b for Male and Female Talkers

						Male Talkers		Female Talkers		All Talkers	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	0.021	0.435	0.042	0.614	0.031	0.531
2	2	Street	AMR@x	-	AMR@x	0.052	0.622	0.021	0.542	0.036	0.582
3	3	Babble	AMR@x	-	AMR@x	0.021	0.696	-0.083	0.610	-0.031	0.654
4	4	Car	MNRU-16	-	MNRU-12	-0.781	0.941	-0.750	0.984	-0.766	0.955
5	5	Car	Direct	-	MNRU-12	-2.406	0.875	-2.344	0.865	-2.375	0.864
6	4'	Street	MNRU-16	-	MNRU-12	-0.719	0.813	-0.781	0.975	-0.750	0.891
7	5'	Street	Direct	-	MNRU-12	-2.031	0.967	-2.406	0.837	-2.219	0.917
8	4''	Babble	MNRU-16	-	MNRU-12	-0.688	0.931	-0.563	1.076	-0.625	1.000
9	5''	Babble	Direct	-	MNRU-12	-2.281	0.888	-2.313	1.203	-2.297	1.049
10	6	Car	AMR@x	4	AMR@x	0.427	0.645	0.188	0.549	0.307	0.609
11	7	Car	AMR@x	7	AMR@x	0.438	0.708	0.594	0.828	0.516	0.772
12	8	Car	AMR@x	10	AMR@x	0.656	0.856	0.740	0.798	0.698	0.827
13	9	Street	AMR@x	4	AMR@x	0.250	0.598	0.188	0.654	0.219	0.626
14	10	Street	AMR@x	7	AMR@x	0.417	0.763	0.521	0.781	0.469	0.772
15	11	Street	AMR@x	10	AMR@x	0.552	0.905	0.719	0.855	0.635	0.882
16	12	Babble	AMR@x	4	AMR@x	0.219	0.684	0.177	0.725	0.198	0.703
17	13	Babble	AMR@x	7	AMR@x	0.313	0.812	0.458	0.857	0.385	0.836
18	14	Babble	AMR@x	10	AMR@x	0.625	0.897	0.458	0.832	0.542	0.867
19	15	Car	AMR@x	-	AMR/NS1@x	0.385	0.887	0.458	0.893	0.422	0.889
20	16	Car	AMR@x	-	AMR/NS2@x	0.542	0.857	0.563	0.868	0.552	0.861
21	17	Car	AMR@x	-	AMR/NS3@x	0.510	0.781	0.438	0.805	0.474	0.792
22	18	Car	AMR@x	-	AMR/NS4@x	0.469	0.858	0.583	0.829	0.526	0.843
23	19	Car	AMR@x	-	AMR/NS5@x	0.615	0.910	0.417	0.925	0.516	0.921
24	20	Car	AMR@x	-	AMR/NS6@x	0.521	1.095	0.604	0.957	0.563	1.026
25	21	Street	AMR@x	-	AMR/NS1@x	0.354	0.833	0.427	0.764	0.391	0.798
26	22	Street	AMR@x	-	AMR/NS2@x	0.469	0.820	0.375	0.849	0.422	0.834
27	23	Street	AMR@x	-	AMR/NS3@x	0.510	0.754	0.354	0.882	0.432	0.822
28	24	Street	AMR@x	-	AMR/NS4@x	0.458	0.845	0.688	0.812	0.573	0.834
29	25	Street	AMR@x	-	AMR/NS5@x	0.208	0.695	0.427	0.843	0.318	0.778
30	26	Street	AMR@x	-	AMR/NS6@x	0.875	0.976	0.750	1.046	0.813	1.011
31	27	Babble	AMR@x	-	AMR/NS1@x	0.240	0.736	0.292	0.695	0.266	0.714
32	28	Babble	AMR@x	-	AMR/NS2@x	0.260	0.729	0.229	0.827	0.245	0.778
33	29	Babble	AMR@x	-	AMR/NS3@x	0.031	0.774	0.240	0.736	0.135	0.761
34	30	Babble	AMR@x	-	AMR/NS4@x	0.281	0.736	0.271	0.718	0.276	0.725
35	31	Babble	AMR@x	-	AMR/NS5@x	0.208	0.679	0.229	0.732	0.219	0.705
36	32	Babble	AMR@x	-	AMR/NS6@x	0.417	0.937	0.260	0.824	0.339	0.883

Table 4b2: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 4b for Normal and Reversed Presentation Orders

						Normal Order		Reversed Order		Both Orders	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	-0.031	0.469	0.094	0.582	0.031	0.531
2	2	Street	AMR@x	-	AMR@x	0.010	0.552	0.063	0.612	0.036	0.582
3	3	Babble	AMR@x	-	AMR@x	0.094	0.617	-0.156	0.670	-0.031	0.654
4	4	Car	MNRU-16	-	MNRU-12	-1.250	0.803	-0.281	0.851	-0.766	0.955
5	5	Car	Direct	-	MNRU-12	-2.531	0.718	-2.219	0.975	-2.375	0.864
6	4'	Street	MNRU-16	-	MNRU-12	-1.094	0.893	-0.406	0.756	-0.750	0.891
7	5'	Street	Direct	-	MNRU-12	-2.250	0.842	-2.188	0.998	-2.219	0.917
8	4"	Babble	MNRU-16	-	MNRU-12	-1.094	0.818	-0.156	0.954	-0.625	1.000
9	5"	Babble	Direct	-	MNRU-12	-2.313	0.738	-2.281	1.301	-2.297	1.049
10	6	Car	AMR@x	4	AMR@x	0.344	0.613	0.271	0.607	0.307	0.609
11	7	Car	AMR@x	7	AMR@x	0.667	0.816	0.365	0.698	0.516	0.772
12	8	Car	AMR@x	10	AMR@x	0.885	0.793	0.510	0.821	0.698	0.827
13	9	Street	AMR@x	4	AMR@x	0.333	0.643	0.104	0.589	0.219	0.626
14	10	Street	AMR@x	7	AMR@x	0.563	0.856	0.375	0.669	0.469	0.772
15	11	Street	AMR@x	10	AMR@x	0.958	0.780	0.313	0.862	0.635	0.882
16	12	Babble	AMR@x	4	AMR@x	0.479	0.711	-0.083	0.574	0.198	0.703
17	13	Babble	AMR@x	7	AMR@x	0.729	0.840	0.042	0.679	0.385	0.836
18	14	Babble	AMR@x	10	AMR@x	0.896	0.864	0.188	0.715	0.542	0.867
19	15	Car	AMR@x	-	AMR/NS1@x	0.448	0.961	0.396	0.814	0.422	0.889
20	16	Car	AMR@x	-	AMR/NS2@x	0.677	0.827	0.427	0.880	0.552	0.861
21	17	Car	AMR@x	-	AMR/NS3@x	0.500	0.795	0.448	0.793	0.474	0.792
22	18	Car	AMR@x	-	AMR/NS4@x	0.646	0.808	0.406	0.865	0.526	0.843
23	19	Car	AMR@x	-	AMR/NS5@x	0.698	0.896	0.333	0.914	0.516	0.921
24	20	Car	AMR@x	-	AMR/NS6@x	0.615	1.080	0.510	0.973	0.563	1.026
25	21	Street	AMR@x	-	AMR/NS1@x	0.479	0.833	0.302	0.756	0.391	0.798
26	22	Street	AMR@x	-	AMR/NS2@x	0.583	0.879	0.260	0.757	0.422	0.834
27	23	Street	AMR@x	-	AMR/NS3@x	0.521	0.858	0.344	0.779	0.432	0.822
28	24	Street	AMR@x	-	AMR/NS4@x	0.865	0.790	0.281	0.777	0.573	0.834
29	25	Street	AMR@x	-	AMR/NS5@x	0.510	0.754	0.125	0.757	0.318	0.778
30	26	Street	AMR@x	-	AMR/NS6@x	1.104	1.051	0.521	0.882	0.813	1.011
31	27	Babble	AMR@x	-	AMR/NS1@x	0.552	0.694	-0.021	0.615	0.266	0.714
32	28	Babble	AMR@x	-	AMR/NS2@x	0.521	0.882	-0.031	0.532	0.245	0.778
33	29	Babble	AMR@x	-	AMR/NS3@x	0.385	0.800	-0.115	0.630	0.135	0.761
34	30	Babble	AMR@x	-	AMR/NS4@x	0.552	0.780	0.000	0.543	0.276	0.725
35	31	Babble	AMR@x	-	AMR/NS5@x	0.500	0.681	-0.063	0.612	0.219	0.705
36	32	Babble	AMR@x	-	AMR/NS6@x	0.719	0.926	-0.042	0.648	0.339	0.883

Experiment 5a

Variances From the Test Plan

Due to the large response differences between normal-order and reverse-order trials and the need to replace subjects in Experiments 4a and 4b, the instructions to subjects were modified for Experiments 5a, 5b, 6 and 7. These modifications were to simplify the instructions by removing mention of a reference sample and a telecommunications system. Also, the new instructions emphasized that the ratings should always be of the second voice sample. It was hoped that this would reduce any chance that subjects were not using the rating scale consistently for all the trials.

The relevant sections of the original instructions were...

> In this experiment we are evaluating systems that might be used for telecommunication services.
>
> You are going to hear through the handset pairs of speech samples, most of which have been recorded in different noisy environments (for example inside a car, in an office, or on the street). The first sample you will hear will be the reference sample. You will then hear the same sample again, but this time it will have passed through a telecommunications system. These samples will each consist of two short unconnected sentences.
>
> You should listen carefully to each pair of samples. When they have finished, please record your opinion of the second sample with regard to the first one using the following scale...

The corresponding sections of the revised instructions were

> Through the headphones, you are going to hear pairs of speech samples, most of which have been recorded in different noisy environments (for example inside a car, in an office, or on the street). These samples will each consist of two short unconnected sentences.
>
> You should listen carefully to each pair of samples. When they have finished, the seven labeled buttons on your response screen will light up. Press the button corresponding to **your opinion of the second sample** in comparison to the first one using the following scale...
>
> **<u>Remember</u>**: your task is to give **your opinion of the second sample in comparison to the first one.**

Observations on the Test Procedure

Replacing subjects due to extreme data. Following the procedure adopted for Experiment 4a, the results from each subject were examined in detail. One subject was identified as having extreme differences between their normal-order and reversed-order trials, and this subject was replaced. Following this replacement, the mean difference between normal and reversed-order trials for Conditions 6-32 in the test plan was 0.59.

Results

Table 5a.1 below summarizes the final results of Experiment 5a in terms of means and standard deviations for each condition, preserving the information about talker gender. Table 5a.2 preserves the information about presentation order. Also, the results of the experiment have been reported in the soft copy of the spreadsheet provided by John Tardelli.

Table 5a.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 5a for Male and Female Talkers

						Male Talkers		Female Talkers		All Talkers	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	0.042	0.541	-0.031	0.623	0.005	0.583
2	2	Street	AMR@x	-	AMR@x	0.063	0.539	0.094	0.666	0.078	0.605
3	3	Babble	AMR@x	-	AMR@x	0.000	0.740	0.021	0.711	0.010	0.723
4	4	Car	MNRU-16	-	MNRU-12	-0.656	1.208	-0.469	0.915	-0.563	1.067
5	5	Car	Direct	-	MNRU-12	-2.094	0.856	-2.219	0.792	-2.156	0.821
6	4'	Street	MNRU-16	-	MNRU-12	-0.750	1.164	-1.125	1.129	-0.938	1.153
7	5'	Street	Direct	-	MNRU-12	-2.313	0.644	-2.469	0.671	-2.391	0.657
8	4"	Babble	MNRU-16	-	MNRU-12	-0.844	0.920	-0.906	0.856	-0.875	0.882
9	5"	Babble	Direct	-	MNRU-12	-2.188	0.738	-2.156	0.723	-2.172	0.725
10	6	Car	AMR@x	4	AMR@x	0.479	0.615	0.302	0.713	0.391	0.670
11	7	Car	AMR@x	7	AMR@x	0.771	0.732	0.823	0.846	0.797	0.790
12	8	Car	AMR@x	10	AMR@x	0.958	0.832	1.083	0.854	1.021	0.844
13	9	Street	AMR@x	4	AMR@x	0.406	0.734	0.406	0.762	0.406	0.746
14	10	Street	AMR@x	7	AMR@x	0.615	0.887	0.677	0.827	0.646	0.856
15	11	Street	AMR@x	10	AMR@x	0.979	0.973	1.073	1.136	1.026	1.056
16	12	Babble	AMR@x	4	AMR@x	0.281	0.791	0.417	0.925	0.349	0.861
17	13	Babble	AMR@x	7	AMR@x	0.531	0.929	0.573	1.003	0.552	0.964
18	14	Babble	AMR@x	10	AMR@x	0.844	1.108	0.833	1.102	0.839	1.102
19	15	Car	AMR@x	-	AMR/NS1@x	0.479	0.929	0.698	0.848	0.589	0.894
20	16	Car	AMR@x	-	AMR/NS2@x	0.646	1.179	0.781	1.107	0.714	1.142
21	17	Car	AMR@x	-	AMR/NS3@x	0.531	0.820	0.385	0.956	0.458	0.891
22	18	Car	AMR@x	-	AMR/NS4@x	0.938	0.916	0.635	0.975	0.786	0.955
23	19	Car	AMR@x	-	AMR/NS5@x	0.833	0.867	0.396	1.021	0.615	0.969
24	20	Car	AMR@x	-	AMR/NS6@x	0.813	1.371	0.552	1.141	0.682	1.265
25	21	Street	AMR@x	-	AMR/NS1@x	0.719	0.959	0.729	0.888	0.724	0.922
26	22	Street	AMR@x	-	AMR/NS2@x	0.750	0.995	0.583	0.981	0.667	0.989
27	23	Street	AMR@x	-	AMR/NS3@x	0.604	0.978	0.563	0.892	0.583	0.934
28	24	Street	AMR@x	-	AMR/NS4@x	0.792	0.939	0.615	0.910	0.703	0.927
29	25	Street	AMR@x	-	AMR/NS5@x	0.240	0.818	0.417	0.902	0.328	0.863
30	26	Street	AMR@x	-	AMR/NS6@x	0.688	1.276	0.927	1.163	0.807	1.223
31	27	Babble	AMR@x	-	AMR/NS1@x	0.625	0.987	0.500	1.026	0.563	1.006
32	28	Babble	AMR@x	-	AMR/NS2@x	0.344	0.950	0.469	0.951	0.406	0.950
33	29	Babble	AMR@x	-	AMR/NS3@x	0.333	0.948	0.302	0.942	0.318	0.942
34	30	Babble	AMR@x	-	AMR/NS4@x	0.510	0.951	0.458	1.045	0.484	0.997
35	31	Babble	AMR@x	-	AMR/NS5@x	0.250	0.894	0.094	0.884	0.172	0.890
36	32	Babble	AMR@x	-	AMR/NS6@x	0.542	1.230	0.552	1.065	0.547	1.148

Table 5a.2: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 5a for Normal and Reversed Presentation Orders

						Normal Order		Reversed Order		Both Orders	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	-0.146	0.562	0.156	0.568	0.005	0.583
2	2	Street	AMR@x	-	AMR@x	0.031	0.589	0.125	0.620	0.078	0.605
3	3	Babble	AMR@x	-	AMR@x	0.302	0.713	-0.281	0.610	0.010	0.723
4	4	Car	MNRU-16	-	MNRU-12	-1.156	1.051	0.031	0.695	-0.563	1.067
5	5	Car	Direct	-	MNRU-12	-2.594	0.499	-1.719	0.851	-2.156	0.821
6	4'	Street	MNRU-16	-	MNRU-12	-1.656	0.902	-0.219	0.906	-0.938	1.153
7	5'	Street	Direct	-	MNRU-12	-2.531	0.621	-2.250	0.672	-2.391	0.657
8	4"	Babble	MNRU-16	-	MNRU-12	-1.313	0.738	-0.438	0.801	-0.875	0.882
9	5"	Babble	Direct	-	MNRU-12	-2.313	0.644	-2.031	0.782	-2.172	0.725
10	6	Car	AMR@x	4	AMR@x	0.354	0.598	0.427	0.736	0.391	0.670
11	7	Car	AMR@x	7	AMR@x	0.917	0.816	0.677	0.747	0.797	0.790
12	8	Car	AMR@x	10	AMR@x	1.260	0.861	0.781	0.757	1.021	0.844
13	9	Street	AMR@x	4	AMR@x	0.573	0.707	0.240	0.750	0.406	0.746
14	10	Street	AMR@x	7	AMR@x	0.969	0.827	0.323	0.761	0.646	0.856
15	11	Street	AMR@x	10	AMR@x	1.479	0.846	0.573	1.054	1.026	1.056
16	12	Babble	AMR@x	4	AMR@x	0.750	0.871	-0.052	0.639	0.349	0.861
17	13	Babble	AMR@x	7	AMR@x	1.073	0.885	0.031	0.732	0.552	0.964
18	14	Babble	AMR@x	10	AMR@x	1.458	1.015	0.219	0.797	0.839	1.102
19	15	Car	AMR@x	-	AMR/NS1@x	0.479	0.940	0.698	0.835	0.589	0.894
20	16	Car	AMR@x	-	AMR/NS2@x	0.781	1.198	0.646	1.086	0.714	1.142
21	17	Car	AMR@x	-	AMR/NS3@x	0.375	0.874	0.542	0.905	0.458	0.891
22	18	Car	AMR@x	-	AMR/NS4@x	0.896	1.021	0.677	0.877	0.786	0.955
23	19	Car	AMR@x	-	AMR/NS5@x	0.740	0.932	0.490	0.995	0.615	0.969
24	20	Car	AMR@x	-	AMR/NS6@x	0.698	1.408	0.667	1.111	0.682	1.265
25	21	Street	AMR@x	-	AMR/NS1@x	1.042	0.928	0.406	0.802	0.724	0.922
26	22	Street	AMR@x	-	AMR/NS2@x	0.969	0.978	0.365	0.908	0.667	0.989
27	23	Street	AMR@x	-	AMR/NS3@x	0.917	0.914	0.250	0.834	0.583	0.934
28	24	Street	AMR@x	-	AMR/NS4@x	1.021	0.940	0.385	0.800	0.703	0.927
29	25	Street	AMR@x	-	AMR/NS5@x	0.510	0.871	0.146	0.821	0.328	0.863
30	26	Street	AMR@x	-	AMR/NS6@x	1.219	1.224	0.396	1.081	0.807	1.223
31	27	Babble	AMR@x	-	AMR/NS1@x	1.146	0.951	-0.021	0.665	0.563	1.006
32	28	Babble	AMR@x	-	AMR/NS2@x	0.823	1.046	-0.010	0.607	0.406	0.950
33	29	Babble	AMR@x	-	AMR/NS3@x	0.771	0.989	-0.135	0.626	0.318	0.942
34	30	Babble	AMR@x	-	AMR/NS4@x	1.010	0.946	-0.042	0.739	0.484	0.997
35	31	Babble	AMR@x	-	AMR/NS5@x	0.677	0.801	-0.333	0.660	0.172	0.890
36	32	Babble	AMR@x	-	AMR/NS6@x	1.167	1.130	-0.073	0.771	0.547	1.148

Experiment 5b

Observations on the Test Procedure

Replacing subjects due to extreme data. Following the procedure adopted for Experiment 4a, the results from each subject were examined in detail. No subjects had to be replaced in this experiment. The mean difference between normal and reversed-order trials for Conditions 6-32 in the test plan was 0.54.

Results

Table 5b.1 below summarizes the results of Experiment 5b in terms of means and standard deviations for each condition, preserving the information about talker gender. Table 5b.2 preserves the information about presentation order. Also, the results of the experiment have been reported in the soft copy of the spreadsheet provided by John Tardelli.

Table 5b.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 5b for Male and Female Talkers

						Male Talkers		Female Talkers		All Talkers	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	0.052	0.622	0.042	0.631	0.047	0.625
2	2	Street	AMR@x	-	AMR@x	-0.042	0.767	-0.052	0.716	-0.047	0.740
3	3	Babble	AMR@x	-	AMR@x	-0.073	0.757	-0.156	0.825	-0.115	0.791
4	4	Car	MNRU-16	-	MNRU-12	-0.875	1.100	-0.906	0.856	-0.891	0.978
5	5	Car	Direct	-	MNRU-12	-2.375	0.907	-2.688	0.535	-2.531	0.755
6	4'	Street	MNRU-16	-	MNRU-12	-0.844	1.081	-0.969	0.967	-0.906	1.019
7	5'	Street	Direct	-	MNRU-12	-2.313	0.780	-2.375	0.660	-2.344	0.718
8	4''	Babble	MNRU-16	-	MNRU-12	-0.531	1.107	-0.625	0.793	-0.578	0.956
9	5''	Babble	Direct	-	MNRU-12	-2.844	0.369	-2.719	0.772	-2.781	0.603
10	6	Car	AMR@x	4	AMR@x	0.417	0.721	0.406	0.689	0.411	0.703
11	7	Car	AMR@x	7	AMR@x	0.729	0.852	0.781	0.784	0.755	0.817
12	8	Car	AMR@x	10	AMR@x	0.938	0.904	1.104	0.774	1.021	0.844
13	9	Street	AMR@x	4	AMR@x	0.323	0.761	0.354	0.821	0.339	0.789
14	10	Street	AMR@x	7	AMR@x	0.646	0.906	0.604	0.814	0.625	0.859
15	11	Street	AMR@x	10	AMR@x	0.875	0.976	1.073	1.039	0.974	1.010
16	12	Babble	AMR@x	4	AMR@x	0.313	0.898	0.448	0.752	0.380	0.829
17	13	Babble	AMR@x	7	AMR@x	0.521	0.882	0.625	0.897	0.573	0.889
18	14	Babble	AMR@x	10	AMR@x	0.792	1.142	0.792	0.917	0.792	1.033
19	15	Car	AMR@x	-	AMR/NS1@x	0.542	1.004	0.677	1.041	0.609	1.023
20	16	Car	AMR@x	-	AMR/NS2@x	0.688	0.955	0.802	0.947	0.745	0.950
21	17	Car	AMR@x	-	AMR/NS3@x	0.646	0.984	0.667	0.879	0.656	0.930
22	18	Car	AMR@x	-	AMR/NS4@x	0.844	0.875	0.979	0.882	0.911	0.879
23	19	Car	AMR@x	-	AMR/NS5@x	0.750	0.906	0.740	0.954	0.745	0.928
24	20	Car	AMR@x	-	AMR/NS6@x	0.708	1.321	0.833	1.237	0.771	1.278
25	21	Street	AMR@x	-	AMR/NS1@x	0.427	1.034	0.531	1.005	0.479	1.018
26	22	Street	AMR@x	-	AMR/NS2@x	0.479	0.984	0.510	0.906	0.495	0.943
27	23	Street	AMR@x	-	AMR/NS3@x	0.521	0.973	0.448	0.961	0.484	0.965
28	24	Street	AMR@x	-	AMR/NS4@x	0.677	0.968	0.729	0.968	0.703	0.966
29	25	Street	AMR@x	-	AMR/NS5@x	0.260	1.008	0.521	0.962	0.391	0.991
30	26	Street	AMR@x	-	AMR/NS6@x	0.948	1.118	0.896	1.156	0.922	1.134
31	27	Babble	AMR@x	-	AMR/NS1@x	0.469	0.906	0.344	1.014	0.406	0.961
32	28	Babble	AMR@x	-	AMR/NS2@x	0.313	0.944	0.344	1.045	0.328	0.993
33	29	Babble	AMR@x	-	AMR/NS3@x	0.219	0.873	0.375	0.861	0.297	0.869
34	30	Babble	AMR@x	-	AMR/NS4@x	0.396	1.021	0.510	0.858	0.453	0.942
35	31	Babble	AMR@x	-	AMR/NS5@x	0.177	0.962	0.104	0.814	0.141	0.890
36	32	Babble	AMR@x	-	AMR/NS6@x	0.417	1.176	0.313	0.944	0.365	1.065

Table 5b.2: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 5b for Normal and Reversed Presentation Orders

						Normal Order		Reversed Order		Both Orders	
True Cond	**Cond**	**Noise**	**Reference**	**Ideal NS**	**Processed**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	1	Car	AMR@x	-	AMR@x	0.042	0.614	0.052	0.639	0.047	0.625
2	2	Street	AMR@x	-	AMR@x	0.083	0.706	-0.177	0.754	-0.047	0.740
3	3	Babble	AMR@x	-	AMR@x	0.208	0.739	-0.438	0.708	-0.115	0.791
4	4	Car	MNRU-16	-	MNRU-12	-1.281	0.958	-0.500	0.842	-0.891	0.978
5	5	Car	Direct	-	MNRU-12	-2.719	0.457	-2.344	0.937	-2.531	0.755
6	4'	Street	MNRU-16	-	MNRU-12	-1.594	0.712	-0.219	0.792	-0.906	1.019
7	5'	Street	Direct	-	MNRU-12	-2.438	0.564	-2.250	0.842	-2.344	0.718
8	4"	Babble	MNRU-16	-	MNRU-12	-0.813	0.998	-0.344	0.865	-0.578	0.956
9	5"	Babble	Direct	-	MNRU-12	-2.813	0.397	-2.750	0.762	-2.781	0.603
10	6	Car	AMR@x	4	AMR@x	0.427	0.722	0.396	0.688	0.411	0.703
11	7	Car	AMR@x	7	AMR@x	0.865	0.776	0.646	0.846	0.755	0.817
12	8	Car	AMR@x	10	AMR@x	1.208	0.807	0.833	0.842	1.021	0.844
13	9	Street	AMR@x	4	AMR@x	0.615	0.745	0.063	0.737	0.339	0.789
14	10	Street	AMR@x	7	AMR@x	0.906	0.769	0.344	0.856	0.625	0.859
15	11	Street	AMR@x	10	AMR@x	1.406	0.789	0.542	1.025	0.974	1.010
16	12	Babble	AMR@x	4	AMR@x	0.792	0.794	-0.031	0.640	0.380	0.829
17	13	Babble	AMR@x	7	AMR@x	1.063	0.779	0.083	0.706	0.573	0.889
18	14	Babble	AMR@x	10	AMR@x	1.375	0.932	0.208	0.767	0.792	1.033
19	15	Car	AMR@x	-	AMR/NS1@x	0.688	1.009	0.531	1.036	0.609	1.023
20	16	Car	AMR@x	-	AMR/NS2@x	0.719	1.033	0.771	0.864	0.745	0.950
21	17	Car	AMR@x	-	AMR/NS3@x	0.708	0.951	0.604	0.912	0.656	0.930
22	18	Car	AMR@x	-	AMR/NS4@x	1.083	0.879	0.740	0.849	0.911	0.879
23	19	Car	AMR@x	-	AMR/NS5@x	0.875	0.897	0.615	0.944	0.745	0.928
24	20	Car	AMR@x	-	AMR/NS6@x	0.823	1.392	0.719	1.158	0.771	1.278
25	21	Street	AMR@x	-	AMR/NS1@x	0.688	1.049	0.271	0.946	0.479	1.018
26	22	Street	AMR@x	-	AMR/NS2@x	0.823	0.894	0.167	0.879	0.495	0.943
27	23	Street	AMR@x	-	AMR/NS3@x	0.760	0.915	0.208	0.939	0.484	0.965
28	24	Street	AMR@x	-	AMR/NS4@x	1.000	0.906	0.406	0.936	0.703	0.966
29	25	Street	AMR@x	-	AMR/NS5@x	0.646	0.984	0.135	0.936	0.391	0.991
30	26	Street	AMR@x	-	AMR/NS6@x	1.229	1.081	0.615	1.109	0.922	1.134
31	27	Babble	AMR@x	-	AMR/NS1@x	0.823	0.940	-0.010	0.788	0.406	0.961
32	28	Babble	AMR@x	-	AMR/NS2@x	0.719	1.093	-0.063	0.693	0.328	0.993
33	29	Babble	AMR@x	-	AMR/NS3@x	0.635	0.896	-0.042	0.695	0.297	0.869
34	30	Babble	AMR@x	-	AMR/NS4@x	0.865	0.969	0.042	0.710	0.453	0.942
35	31	Babble	AMR@x	-	AMR/NS5@x	0.563	0.916	-0.281	0.627	0.141	0.890
36	32	Babble	AMR@x	-	AMR/NS6@x	0.740	1.225	-0.010	0.703	0.365	1.065

Experiment 6

Observations on the Test Procedure

Error in speech used for practice trials. An error was made in preparing the preliminary (practice) trials for this experiment. The error was caused by a technician failing to replace the "c" in the speech sample filenames with the practice designator "p". The error was in the practice trials only. For all groups, some test trial samples were presented instead of the practice samples as intended in the test plan. All groups used the same order and the same files for the practice. The actual files played were as follows:

Files intended to be played (per test plan)...

M1S01p01	M1S01p09	practice trial 1
M2S01p10	M2S01p02	practice trial 2
F1S01p11	F1S01p03	practice trial 3
F2S01p04	F2S01p12	practice trial 4
M2S01p13	M2S01p05	practice trial 5
M1S01p06	M1S01p14	practice trial 6
F1S01p07	F1S01p15	practice trial 7
F2S01p16	F2S01p08	practice trial 8

Files actually played...

M1S01c01	M1S01c09	trial 136, group 4
M2S01c10	M2S01c02	trial 048, group 1
F1S01c11	F1S01c03	trial 015, group 1
F2S01c04	F2S01c12	trial 013, group 4
M2S01c13	M2S01c05	trial 048, group 6
M1S01c06	M1S01c14	trial 038, group 4
F1S01c07	F1S01c15	trial 067, group 1
F2S01c16	F2S01c08	trial 039, group 6

We do not believe that this error has affected the results in any significant way. This is because the speech samples still represented a good sampling of the speech samples that would be heard during the experiment proper. Moreover, for 3 of the groups the practice samples were repeated during the experiment, whereas for the other 3 groups the practice samples were never heard again.

Replacing subjects due to extreme data. Following the procedure adopted for Experiment 4a, the results from each subject were examined in detail. One subject had to be replaced in this experiment because of extreme differences between normal and reversed-order trials. Following the replacement, the mean difference between normal and reversed-order trials for Conditions 7-24 in the test plan was 0.59.

Results

Table 6.1 below summarizes the results of Experiment 6 in terms of means and standard deviations for each condition, preserving the information about talker gender. Table 6.2 preserves the information about presentation order. Also, the results of the experiment have been reported in the soft copy of the spreadsheet provided by John Tardelli.

Table 6.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 6 for Male and Female Talkers

	Reference		Processed		Male Talkers		Female Talkers		All Talkers	
Condition	**UL**	**DL**	**UL**	**DL**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	Direct	-	Direct	-	0.052	0.587	0.010	0.513	0.031	0.550
2	AMR@x	-	AMR@x	-	0.031	0.640	-0.042	0.521	-0.005	0.583
3	MNRU-16	-	MNRU-12	-	-0.552	0.793	-0.510	0.768	-0.531	0.779
4	Direct	-	MNRU-12	-	-1.896	0.864	-1.865	0.890	-1.880	0.875
5	MNRU-Q1	-	MNRU-12	-	-0.667	0.829	-0.677	0.888	-0.672	0.857
6	MNRU-Q2	-	MNRU-12	-	-0.833	0.829	-0.865	0.902	-0.849	0.864
7	AMR@x	-	AMR/NS1@x	-	0.729	0.989	0.594	0.947	0.661	0.968
8	AMR@x	-	AMR/NS2@x	-	0.781	1.028	0.573	1.149	0.677	1.093
9	AMR@x	-	AMR/NS3@x	-	0.469	0.940	0.406	0.913	0.438	0.925
10	AMR@x	-	AMR/NS4@x	-	0.583	1.063	0.500	1.076	0.542	1.068
11	AMR@x	-	AMR/NS5@x	-	0.594	1.120	0.583	1.043	0.589	1.079
12	AMR@x	-	AMR/NS6@x	-	0.885	1.195	0.698	1.134	0.792	1.166
13	AMR@x	AMR@y	AMR/NS1@x	AMR/NS1@y	0.552	1.329	0.469	1.322	0.510	1.322
14	AMR@x	AMR@y	AMR/NS2@x	AMR/NS2@y	0.563	1.413	0.604	1.183	0.583	1.300
15	AMR@x	AMR@y	AMR/NS3@x	AMR/NS3@y	0.646	1.289	0.448	1.195	0.547	1.244
16	AMR@x	AMR@y	AMR/NS4@x	AMR/NS4@y	0.635	1.331	0.313	1.276	0.474	1.310
17	AMR@x	AMR@y	AMR/NS5@x	AMR/NS5@y	0.708	1.289	0.604	1.252	0.656	1.268
18	AMR@x	AMR@y	AMR/NS6@x	AMR/NS6@y	0.885	1.582	0.563	1.541	0.724	1.566
19	AMR@x	AMR@y	AMR@x	AMR/NS1@y	0.250	0.906	0.292	0.928	0.271	0.915
20	AMR@x	AMR@y	AMR@x	AMR/NS2@y	0.396	1.128	0.469	0.940	0.432	1.036
21	AMR@x	AMR@y	AMR@x	AMR/NS3@y	0.281	0.948	0.396	0.888	0.339	0.918
22	AMR@x	AMR@y	AMR@x	AMR/NS4@y	0.521	0.846	0.396	1.119	0.458	0.991
23	AMR@x	AMR@y	AMR@x	AMR/NS5@y	0.385	0.899	0.427	0.971	0.406	0.933
24	AMR@x	AMR@y	AMR@x	AMR/NS6@y	0.427	1.212	0.396	1.227	0.411	1.216

Table 6.2: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 6 for Normal and Reversed Presentation Orders

	Reference		Processed		Normal Order		Reversed Order		Both Orders	
Condition	**UL**	**DL**	**UL**	**DL**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	Direct	-	Direct	-	0.146	0.542	-0.083	0.536	0.031	0.550
2	AMR@x	-	AMR@x	-	0.115	0.560	-0.125	0.585	-0.005	0.583
3	MNRU-16	-	MNRU-12	-	-0.927	0.743	-0.135	0.592	-0.531	0.779
4	Direct	-	MNRU-12	-	-1.979	0.767	-1.781	0.965	-1.880	0.875
5	MNRU-Q1	-	MNRU-12	-	-1.094	0.847	-0.250	0.632	-0.672	0.857
6	MNRU-Q2	-	MNRU-12	-	-1.281	0.830	-0.417	0.660	-0.849	0.864
7	AMR@x	-	AMR/NS1@x	-	0.979	0.894	0.344	0.938	0.661	0.968
8	AMR@x	-	AMR/NS2@x	-	1.083	1.033	0.271	1.000	0.677	1.093
9	AMR@x	-	AMR/NS3@x	-	0.760	0.791	0.115	0.939	0.438	0.925
10	AMR@x	-	AMR/NS4@x	-	0.760	0.981	0.323	1.110	0.542	1.068
11	AMR@x	-	AMR/NS5@x	-	1.031	0.989	0.146	0.984	0.589	1.079
12	AMR@x	-	AMR/NS6@x	-	1.208	1.075	0.375	1.107	0.792	1.166
13	AMR@x	AMR@y	AMR/NS1@x	AMR/NS1@y	0.677	1.497	0.344	1.103	0.510	1.322
14	AMR@x	AMR@y	AMR/NS2@x	AMR/NS2@y	0.802	1.396	0.365	1.162	0.583	1.300
15	AMR@x	AMR@y	AMR/NS3@x	AMR/NS3@y	0.813	1.284	0.281	1.149	0.547	1.244
16	AMR@x	AMR@y	AMR/NS4@x	AMR/NS4@y	0.854	1.392	0.094	1.106	0.474	1.310
17	AMR@x	AMR@y	AMR/NS5@x	AMR/NS5@y	1.063	1.177	0.250	1.231	0.656	1.268
18	AMR@x	AMR@y	AMR/NS6@x	AMR/NS6@y	1.219	1.571	0.229	1.403	0.724	1.566
19	AMR@x	AMR@y	AMR@x	AMR/NS1@y	0.313	0.933	0.229	0.900	0.271	0.915
20	AMR@x	AMR@y	AMR@x	AMR/NS2@y	0.677	1.129	0.188	0.874	0.432	1.036
21	AMR@x	AMR@y	AMR@x	AMR/NS3@y	0.510	0.962	0.167	0.842	0.339	0.918
22	AMR@x	AMR@y	AMR@x	AMR/NS4@y	0.792	0.939	0.125	0.932	0.458	0.991
23	AMR@x	AMR@y	AMR@x	AMR/NS5@y	0.573	0.981	0.240	0.855	0.406	0.933
24	AMR@x	AMR@y	AMR@x	AMR/NS6@y	0.740	1.283	0.083	1.053	0.411	1.216

Experiment 7

Observations on the Test Procedure

Replacing subjects due to extreme data. Following the procedure adopted for Experiment 4a, the results from each subject were examined in detail. One subject had to be replaced in this experiment because of extreme differences between normal and reversed-order trials. Following the replacement, the mean difference between normal and reversed-order trials for Conditions 7-24 in the test plan was 0.09.

Results

Table 7.1 below summarizes the results of Experiment 7 in terms of means and standard deviations for each condition, preserving the information about talker gender. Table 7.2 preserves the information about presentation order. Also, the results of the experiment have been reported in the soft copy of the spreadsheet provided by John Tardelli.

Table 7.1: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 7 for Male and Female Talkers

	Reference		Processed		Male Talkers		Female Talkers		All Talkers	
Condition	**UL**	**DL**	**UL**	**DL**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	Direct	-	Direct	-	0.031	0.876	0.083	0.879	0.057	0.875
2	AMR@x	-	AMR@x	-	-0.031	0.732	0.042	0.794	0.005	0.762
3	MNRU-16	-	MNRU-12	-	-0.625	0.954	-0.615	1.050	-0.620	1.001
4	Direct	-	MNRU-12	-	-1.625	1.145	-1.781	1.078	-1.703	1.112
5	MNRU-Q1	-	MNRU-12	-	-0.656	1.177	-0.833	0.991	-0.745	1.089
6	MNRU-Q2	-	MNRU-12	-	-0.927	1.136	-0.969	1.051	-0.948	1.091
7	AMR@x	-	AMR/NS1@x	-	0.354	1.114	0.167	1.023	0.260	1.071
8	AMR@x	-	AMR/NS2@x	-	0.458	0.939	0.052	1.019	0.255	0.999
9	AMR@x	-	AMR/NS3@x	-	0.156	0.955	0.146	1.015	0.151	0.983
10	AMR@x	-	AMR/NS4@x	-	0.375	1.117	0.135	1.082	0.255	1.103
11	AMR@x	-	AMR/NS5@x	-	0.292	0.882	0.240	1.013	0.266	0.947
12	AMR@x	-	AMR/NS6@x	-	0.656	1.238	0.615	1.372	0.635	1.303
13	AMR@x	AMR@y	AMR/NS1@x	AMR/NS1@y	0.063	1.457	-0.073	1.324	-0.005	1.390
14	AMR@x	AMR@y	AMR/NS2@x	AMR/NS2@y	-0.146	1.265	-0.198	1.319	-0.172	1.289
15	AMR@x	AMR@y	AMR/NS3@x	AMR/NS3@y	0.406	1.286	0.219	1.163	0.313	1.226
16	AMR@x	AMR@y	AMR/NS4@x	AMR/NS4@y	0.271	1.334	-0.094	1.346	0.089	1.349
17	AMR@x	AMR@y	AMR/NS5@x	AMR/NS5@y	0.333	1.245	0.042	1.196	0.188	1.226
18	AMR@x	AMR@y	AMR/NS6@x	AMR/NS6@y	0.260	1.598	0.229	1.573	0.245	1.581
19	AMR@x	AMR@y	AMR@x	AMR/NS1@y	0.135	1.166	0.417	0.981	0.276	1.084
20	AMR@x	AMR@y	AMR@x	AMR/NS2@y	0.000	1.124	0.063	0.880	0.031	1.007
21	AMR@x	AMR@y	AMR@x	AMR/NS3@y	0.260	1.028	0.260	0.920	0.260	0.973
22	AMR@x	AMR@y	AMR@x	AMR/NS4@y	0.021	1.124	0.250	1.046	0.135	1.089
23	AMR@x	AMR@y	AMR@x	AMR/NS5@y	0.146	0.870	0.156	0.772	0.151	0.821
24	AMR@x	AMR@y	AMR@x	AMR/NS6@y	0.375	1.308	0.365	1.125	0.370	1.217

Table 7.2: Mean Opinion Scores (and Standard Deviations) For Each Condition in Experiment 7 for Normal and Reversed Presentation Orders

	Reference		Processed		Normal Order		Reversed Order		Both Orders	
Condition	**UL**	**DL**	**UL**	**DL**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1	Direct	-	Direct	-	0.260	0.798	-0.146	0.906	0.057	0.875
2	AMR@x	-	AMR@x	-	0.083	0.777	-0.073	0.743	0.005	0.762
3	MNRU-16	-	MNRU-12	-	-1.104	0.946	-0.135	0.803	-0.620	1.001
4	Direct	-	MNRU-12	-	-1.813	1.079	-1.594	1.139	-1.703	1.112
5	MNRU-Q1	-	MNRU-12	-	-1.219	1.028	-0.271	0.934	-0.745	1.089
6	MNRU-Q2	-	MNRU-12	-	-1.396	1.031	-0.500	0.962	-0.948	1.091
7	AMR@x	-	AMR/NS1@x	-	0.417	1.002	0.104	1.119	0.260	1.071
8	AMR@x	-	AMR/NS2@x	-	0.438	1.074	0.073	0.885	0.255	0.999
9	AMR@x	-	AMR/NS3@x	-	0.271	1.031	0.031	0.923	0.151	0.983
10	AMR@x	-	AMR/NS4@x	-	0.417	1.158	0.094	1.027	0.255	1.103
11	AMR@x	-	AMR/NS5@x	-	0.427	0.949	0.104	0.923	0.266	0.947
12	AMR@x	-	AMR/NS6@x	-	1.000	1.257	0.271	1.252	0.635	1.303
13	AMR@x	AMR@y	AMR/NS1@x	AMR/NS1@y	-0.104	1.546	0.094	1.215	-0.005	1.390
14	AMR@x	AMR@y	AMR/NS2@x	AMR/NS2@y	-0.313	1.503	-0.031	1.020	-0.172	1.289
15	AMR@x	AMR@y	AMR/NS3@x	AMR/NS3@y	0.167	1.343	0.458	1.085	0.313	1.226
16	AMR@x	AMR@y	AMR/NS4@x	AMR/NS4@y	0.031	1.532	0.146	1.142	0.089	1.349
17	AMR@x	AMR@y	AMR/NS5@x	AMR/NS5@y	0.063	1.360	0.313	1.069	0.188	1.226
18	AMR@x	AMR@y	AMR/NS6@x	AMR/NS6@y	0.292	1.829	0.198	1.295	0.245	1.581
19	AMR@x	AMR@y	AMR@x	AMR/NS1@y	0.344	1.168	0.208	0.994	0.276	1.084
20	AMR@x	AMR@y	AMR@x	AMR/NS2@y	-0.052	1.182	0.115	0.793	0.031	1.007
21	AMR@x	AMR@y	AMR@x	AMR/NS3@y	0.344	1.094	0.177	0.833	0.260	0.973
22	AMR@x	AMR@y	AMR@x	AMR/NS4@y	0.073	1.181	0.198	0.991	0.135	1.089
23	AMR@x	AMR@y	AMR@x	AMR/NS5@y	0.156	0.921	0.146	0.711	0.151	0.821
24	AMR@x	AMR@y	AMR@x	AMR/NS6@y	0.563	1.352	0.177	1.036	0.370	1.217

Summary Observations

The CCR experiments represented a fairly difficult task for the subjects. Since the reference conditions were coded speech without any noise suppression, it was difficult for subjects to know what criteria to use when rating the processed conditions with noise suppression. Listening to the sound files, there were at least four relevant dimensions on which the reference and processed samples could differ:

1. loudness of the noise signal
2. loudness of the voice signal
3. distortion of the noise signal
4. distortion of the voice signal

It is difficult to know which of these factors were used when making comparison judgments, and whether this was consistent during the course of the experimental sessions. Also, it is likely that different subjects used different strategies and criteria when making their ratings, adding to the variability in the data. In addition, the ratings may be complex, for example when processing results in a reduced noise signal accompanied by distortion in the voice and noise. Without more explicit instructions, subjects don't know how to rate such complex differences. These multiple dimensions may also explain the large differences between normal and reversed-order trials: different dimensions become salient depending on the presentation order.

One change that might improve the experiments is to use reference conditions where an ideal noise suppressor provided the same approximate suppression level as the candidate algorithms. For example, compare each candidate with ideal suppression at 7 dB. It is also a good practice to include a well-known noise suppressor (such as the one used in EVRC) as a reference condition.

A different rating task may have been more informative. For example, a "listening effort" scale might be more appropriate than an overall quality comparison. Here subjects would be asked to rate the reference and processed speech in terms of the effort required to understand the voice sample among the background noise using the following scale:

- Complete relaxation possible; no effort required
- Attention necessary; no appreciable effort required
- Moderate effort required
- Considerable effort required
- No meaning understood with any feasible effort

Antony Rix of British Telecom recently described the advantages and disadvantages of various subjective rating scales for experiments with background noise (British Telecom (U.K.), " Comparison Of Opinion Scales For Subjective Listening Tests". ITU-T SG12, Delayed Contribution 102, Sept. 21-30, 1999; Geneva. Contact: A. Rix), and this should be considered when planning any future work.