TNO Human Factors

Kampweg 5
P.O. Box 23
3769 ZG SOESTERBERG
The Netherlands

www.tno.nl

T +31 346 356 211
F +31 346 353 977

TNO-report

TM-01-C037

NATO-NBVC 1200/2400 bps coder selection; results of TNO tests in phase 2

Date	October 31, 2001
Author(s)	S.J. van Wijngaarden and R.A. van Buuren

Copy no.	
Number of copies	
Number of pages	36
Number of appendices	1
Contractor	
Project number	70176

CONTENTS Page

Rep.No. TM-01-C037

TNO Human Factors
Soesterberg, The Netherlands

NATO-NBVC 1200/2400 bps coder selection; results of TNO tests in phase 2

S.J. van Wijngaarden and R.A. van Buuren

SUMMARY

Purpose: The NATO Ad-Hoc Working Group on Narrow Band Voice Coding (AHWG-NBVC) was tasked to select a new standard 1200 and 2400 bps vocoder for use throughout NATO. For this purpose, the AHWG has initiated a selection process, that is aimed at evaluating which of three candidates is the most suitable coder.

Method: During the second of three phases, several speech performance tests were carried out by three different test laboratories. TNO Human Factors was one of these laboratories. In phase 2, TNO-HF performed speech intelligibility and speech quality tests, as well a tests which quantify the intelligibility of whispered speech, talker recognizability and language dependency. Tests were carried out under several conditions, including acoustic noise conditions, bit error conditions and tandem conditions.

Results and conclusions: All tests yielded clear quantitative results, which have adequate statistical resolution to draw conclusions regarding differences in performance between coders. The best overall performance was offered by the US MELPe vocoder suite.

1 INTRODUCTION

In 1997, the NATO Ad-Hoc Working Group AC/322(SC/6-AHWG/3) on Narrow Band Voice Coding (AHWG-NBVC) was tasked to select a 1200/2400 bps voice coder for use throughout NATO. This selection was decided to take place through a competition between suitable existing voice coding algorithms. The winning coder will be described in STANAG 4591, which will be drafted by the AHWG-NBVC. This STANAG will hopefully find application in many (strategic and tactical) applications, contributing to end-to-end secure voice communications within and between NATO member nations.
The selection procedure consists of three phases. Phase 1 comprised speech intelligibility and speech quality tests. Phase 2 also features intelligibility and quality tests, in a greater number of conditions than phase 1. Some additional performance characteristics are also tested in phase 2: intelligibility of whispered speech, language dependency and talker recognizability. In phase 3 the coder performance in real-time will be evaluated, by carrying out communicability tests in several scenarios.
In phase 2, tests are carried out by three different test labs. In this report test results from the five different types of tests carried out by TNO Human Factors in phase 2 are given. The five test types are:

- Speech intelligibility: Consonant-Vowel-Consonant (CVC)
- Speech quality: Mean Opinion Score (MOS)
- Intelligibility of whispered speech: Speech Reception Threshold (SRT)
- Talker recognizability: TNO talker recognizability test
- Language dependency: TNO language dependency test based on SRT.

Test results from all different test types, and from all co-operating test laboratories, will be combined to yield an overall performance index for each competing vocoder. The procedure used for obtaining these performance indices has become known as the precision-weighted data combination procedure. This procedure takes differences in precision of the applied test methods into account, and ensures that only statistically significant differences in vocoder performance may lead to differences in the corresponding performance index. The results of the five test types carried out by TNO Human Factors are presented in this report in such a way that they can be incorporated directly into the precision-weighted data combination procedure.
In addition to the data needed to calculate performance indices, additional data is included in this report that may be of interest to the developers of the competing vocoders, or of some general scientific interest.

2 CODERS, CONDITIONS AND SUBJECTS

There are three candidate coder pairs (each candidate can operate at both 1200 and 2400 bps) competing in phase 2. Combined with three reference coders, a total of 9 coders are tested. The procedure for obtaining test stimuli for each of these 9 coders involves the activities of the host lab. For phase 2, the host lab is the NATO Command, Control and Consultation Agency (NC3A). The host lab was provided with unprocessed speech material. In the conditions

featuring acoustic noise, speech and noise were mixed electronically (at a predefined speech-to-noise ratio) before sending the material to the host lab for further processing. This mixing was carried out using TNO proprietary Matlab® scripts. All signal-to-noise ratios (SNR) were based on A-weighted levels, calculated using the SAM-SLM procedure (Van Velden, 1991).
The conditions for the different test types are given in table 1. For the language dependency and whispered speech tests, the only condition is the “clear” condition without noise or any kind of treatment other than the tested vocoders. These tests are therefore omitted from Table 1.

Table 1 Description of the noise conditions for phase 2.

Noise description	SNR (dB)	Operation	Cond. Nr. CVC	Cond. Nr. MOS	Cond. Nr. Talker Recogn.
No noise	–	–	01	01	01
Speech babble	+6	–	02	02	–
Speech babble	+12	–	03	03	–
Modern office	+20	–	04	04	–
MCE Field shelter	+20	–	05	05	–
HMMWV	+6	–	06	06	02
M2 Bradley	+6	–	07	07	–
UH60 BlackHawk	+12	–	08	08	–
F15	+20	–	09	–	–
Volvo	+12	–	10	–	–
1% random bit errors	–	Bit errors	11	09	–
Tandem with CVSD	–	CVSD->coder	12	10	–

The actual processing through the voice coders (including bit error and tandem operations) was done by the host lab. All processed speech material was returned to the test labs after being subjected to a blinding procedure. Test results as given in this report will be referred using specific names. However, during the test procedure all coders were only known as coders 1 through 9. The actual mapping of the candidate and reference coders to these numbers is referred to as the ‘blinding key’; none of the employees or test subjects working at TNO Human Factors were given this key. Hence, the preparation, testing and evaluation of test results (including the writing of this report) could in no way be influenced by any bias towards any of the candidate coders. The ‘deblinding’ of coders was only applied to this report after discussion of the data in the NATO NBVC ad-hoc working group, at which occasion the deblinding key was made known.
All test subjects participating in tests described in this report, with the exception of the talker recognizability test, were university students of various disciplines, between 17 and 24 years of age. They were screened audiologically by means of pure-tone audiograms (125–8000 Hz). The rejection criterion was set at a best-ear maximum hearing loss of 25 dB at any frequency.
The coders, conditions and procedures used for processing and preparation of test conditions are described in more detail in the test plan (Tardelli *et al.*, 2000).

3 TEST METHODS

3.1 CVC method

The preferred method for measuring speech intelligibility of voice coders at TNO Human Factors is the CVC-method (Steeneken, Geurtsen & Agterhuis, 1990). This method uses simple Consonant-Vowel-Consonant nonsense words; such words, embedded in carrier phrases, were pre-recorded on DAT tape under good laboratory conditions (high quality microphones, no ambient noise). The recorded material consists of speech by eight speakers (4 males and 4 females). Sequences of CVC test-words were combined to obtain word lists of 51 words each. The source material was digitally transferred to a computer mass storage unit by means of an Ariel DAT-link digital audio interface. The speech was resampled to 8 kHz, and stored with 16 bit resolution. This material was mixed electronically with the appropriate acoustic noise files at the SNR levels given in table 1, and sent to the host lab on the designated media. After processing and blinding, the host lab returned the speech material to TNO Human Factors for evaluation.

The processed lists were all presented to a listening panel of 4 listeners, who were asked to respond with the CVC-words as they perceived them. Hence, each data point consists of 32 speaker-listener pairs (8 speakers times 4 listeners). When considering speaker gender as a separate variable, each data point consists of 16 speaker-listener pairs.

Before the test sessions, the subjects were trained. The actual responses of the subjects were given by typing the perceived CVC-words on computer terminals, connected to a central server computer.

As described in section 2, all coders were tested in various acoustic noise conditions. The presentation sequence was balanced for coders, conditions and speaker gender, to minimise the possibility of biased results due to learning effects.

All CVC word scores given in this report are so-called “equally balanced” CVC word scores. Since all phonemes have the same frequency of occurrence in the corpus of test stimuli, the CVC word score is by definition equally balanced. The mean equally balanced CVC word score is calculated for each condition, together with its estimated standard error (s.e.). This standard error is a measure for the accuracy of the mean CVC word score. The standard error is related to the standard deviation according to

$$s.e. = \frac{s}{\sqrt{n}} \qquad \textbf{(1)}$$

where s is the standard deviation, n the number of measuring points and $s.e.$ the standard error. The standard deviation is a measure for the spread of the individual measurements; by using Student’s t-distribution, 95% confidence intervals (as required by the test plan) are easily computed from the standard errors.

The CVC test yields, besides the percentage of correctly responded CVC-words, separate scores for initial and final consonants and for vowels. Moreover, confusion matrices are generated, which may be used to gain insight into the detailed performance of a particular coder.

Because of the large number of conditions, it was not feasible to carry out the entire test with a single panel of 4 listeners. Instead, the experiment was split into two parts, in which two separate panels of 4 listeners participated. Each panel was presented with speech from 4 out of a total of 8 speakers, making sure that between-listener variance can not influence the differences in CVC-scores between coders.

3.2 MOS method

To evaluate speech quality, a choice must be made from a host of available test methods. One such method is known as the Mean Opinion Score (MOS) method. With this straightforward method, fragments of speech are presented to the subjects. After each presentation, subjects are asked to rate the quality of the fragment on a 5 point scale (1 indicating ‘bad’; 5 indicating ‘excellent’). Before the beginning of the actual test, subjects were give an operational definition of speech quality and trained using test stimuli varying widely in quality.

The stimuli were taken from a corpus of read sentences that was recorded at the same time as the CVC stimuli (Steeneken and Geurtsen, 1990), consisting of 65 Dutch everyday sentences, each 8 or 9 syllables in length, by the same 8 speakers as the CVC material.

Each MOS stimulus was a concatenation of two sentences, separated by a silence of approximately one second. After presentation of both sentences, the subjects were prompted by means of a short 1 kHz tone (duration 250 ms) to enter their response on a computer terminal. On this terminal, only the numerical keys 1-5 were activated. After a 3 second response period (within which subjects were free to change their response) the response was stored by a central server computer.

To enable a more or less test-independent interpretation of MOS responses, a calibration system is needed. The subjects’ opinion on a stimulus is much influenced by the context in which it is presented; a calibration system provides a standardised context.

The calibration system used in this case is the MNRU (Modulated Noise Reference Unit, ITU-T Recommendation P.810) procedure. This procedure adds a controlled amount of speech-correlated noise at a chosen SNR, usually given the unit dBQ. The MNRU stimuli provide the subject with a standardised context of calibration stimuli having a spread in quality that should cover the complete 5-point MOS scale. For the NBVC phase 2 MOS test, 8 MNRU conditions were included (5-40 dBQ). Unprocessed ‘clean’ speech stimuli were also included in the presentation schedule.

Usually, 16 subjects take part in each MOS experiment, 4 subjects at a time. Each measuring point is normally based on 128 (8*16) speaker-listener pairs. For practical reasons, a few ‘standby’ subjects are always recruited to fill up gaps in the test program. Since no gaps in the program occurred, we were able to include MOS results for 3 ‘standby’ subjects as well as the regular 16 subjects, giving a total of 19 subjects (and a slightly better accuracy). Each measuring point is now based on 152 (8*19) speaker-listener pairs. All subjects were presented with 792 stimuli (9 coders * 10 conditions * 8 speakers gives 720 test stimuli; 8 MNRU and 1 ‘clean’ stimulus * 8 speakers gives 72 stimuli).

3.3 SRT method for whispered speech

Since whispering affects speech intelligibility on a segmental as well as on a supra-segmental level (distortions of prosody), an intelligibility test based on sentences is more appropriate than a phoneme-based test. By using sentences, any interactions between prosodic effects of whispering and vocoder performance will be included in the test results.

A widely used implementation of an intelligibility test method based on sentences is the Speech Reception Threshold method (SRT; Plomp & Mimpen, 1979). Using the SRT method, the intelligibility of sentences is expressed as the Speech-to-Noise ratio (SNR), at which 50% of the sentences from a well-defined set can be correctly repeated by a listener.

The SRT-method makes use of recordings of short, redundant sentences (8 or 9 syllables), read aloud (or in this case whispered) by instructed speakers. These recordings are processed through each of the vocoders, divided into lists of 13 sentences and calibrated. Please note that *no* noise is added to the speech *before* being presented to the vocoders.

For the actual test, suitable listeners are recruited. After an introduction to the test procedure, each listener is taken to a silent room where the tests take place. A set of headphones with acceptably low distortion is used to present the recorded sentences (processed through the vocoders) to the participants. Noise is now added to the processed sentences in order to obtain the desired speech-to-noise ratio. The noise has a spectrum identical to the long-term (whispered) speech spectrum of the speaker, as reproduced by the vocoder under test. This means that the noise masks each frequency band of speech to the same degree. The masking noise is *only* used to manipulate the overall intelligibility of the speech, in order to be able to define a meaningful intelligibility threshold. Since the noise is added after processing, any noise-preprocessing capabilities of the tested vocoders will not affect the speech-to-noise ratio.

After each sentence, the listener responds with the sentence, as he or she perceived it. A test leader compares this response with the actual sentence. After each correct response, the noise level is increased by 2 dB; after each incorrect response, the noise level is decreased by 2 dB. The first sentence is repeated until it is responded correctly, using 4 dB steps. After the first sentence, each sentence is presented only once. From the last 10 presentation levels (or speech-to-noise ratios) the mean is calculated, as well as the standard deviation. This average SNR is what is commonly referred to as the SRT.

Better vocoder performance will result in a better intelligibility of the processed speech. More intelligible speech can tolerate higher noise levels before being reduced to 50% sentence intelligibility. Hence, better vocoders will produce lower SRTs.

In order to be able to use the SRT scores in the precision weighted combination procedure (which combines results of many different tests into a single performance index for each coder), the ‘raw’ SRT scores need to be subjected to some simple transformations.

After obtaining ‘raw’ SRT scores (speech-to-noise ratio corresponding to 50% sentence intelligibility), these SRT scores are normalised in two ways. First, from each SRT score the mean SRT across all scores for the given subject is subtracted. This reduces the effect of inter-subject variability (aptness at the SRT task) on the statistical power of the test. Next, the same type of normalisation is applied regarding talker and sentence-list number. This is a single operation, since talker and sentence lists are only used in fixed combinations. These normalisations do not affect the absolute differences in mean SRT scores between coders; they merely affect the standard errors associated with these means, by eliminating sources of variance from the data that are not of interest for the selection procedure. The effect of these normalisations on the analysis of statistical significance of differences between coders is similar to the application of more complicated ANOVA (ANalysis Of VAriance)-designs. The advantage of this approach over ANOVA is that a certain consistency is maintained with the approach adopted for the language dependency test (see section 3.5), and the ability to use simpler and more straightforward methods of calculating statistical confidence intervals.

Finally, the negative of the actual SRT score is taken. This converts the result from a speech-to-noise ratio to a 'noise-to-speech' ratio, breaking regular conventions of presenting these ratios. The reason for doing this is that the precision weighted combination procedure requires that higher values indicate better performance, instead of vice versa (which is the case with conventional SRT scores).

It should be noted that the SRT is generally a coarser measure of speech intelligibility than (for instance) CVC or DRT. For the assessment of vocoders, phoneme-based methods are generally preferred over the SRT method, unless the SRT has some special advantage (the inclusion of the effects of prosody in the case of whispered speech). The lesser precision of the SRT-method is reflected in the precision-weighting process used for the combination of all measurement results, which will assign a lesser influence to the SRT results. This is a natural (and justified) effect of the precision-weighted combination procedure.

3.4 Language dependency test method

Testing language dependency of voice coding systems requires a universal speech intelligibility test method, available in several languages. The testing paradigm must be identical across languages, and a sufficiently large multi-lingual speech database is needed. Issues such as phonetic balancing and dealing with differences in phoneme inventories for different languages must be addressed.

Most practical difficulties associated with designing a multi-lingual test method are avoided by using sentences instead of phonemes as test tokens. By using sentences from the same domain (short, redundant everyday sentences) matching certain criteria regarding length (7 to 9 syllables) and complexity, similar test implementations are easily obtained in various languages. Issues regarding phonetic balancing are dealt with implicitly – if the sentences are sufficiently representative of everyday conversation, the phonetic content of the sentences will automatically be representative as well. The SRT method is very suitable as a basis of a multi-lingual intelligibility test (Van Wijngaarden, 2001b).

To verify the baseline intelligibility in different languages, SRT tests are carried out with native listeners of these languages. For the NBVC selection procedure, SRT results have been obtained in four languages (English, German, French and Dutch) with four speakers and four listeners for each language. By comparing the baseline SRT score (no vocoder) in a given language with the SRT score of speech processed through a vocoder, the decrease of intelligibility due to the vocoder is quantified. The obtained measure of intelligibility could be applied in similar ways as CVC-results, although usually with somewhat lower precision.

After SRT results have been obtained for all vocoders in all of the different languages, language dependency of a specific vocoder will become apparent as differences in SRT scores across languages. The TNO language dependency test transforms these differences into a single language dependency metric for each vocoder (Van Wijngaarden, 2001a).

First, the raw SRT scores are normalised to reduce inter-subject and inter-talker differences, which are not of interest for our current purpose. Next, the mean SRT score for *unprocessed speech* in the corresponding language is subtracted from the individual SRT scores for all of the vocoders. This eliminates inter-language differences in the implementation of the SRT test.

Finally, the mean SRT value is calculated for each coder-language combination (across speakers and listeners). We will call this mean $M_{i,j}$ where i is the index for coder and j for language. We assume that SRT results have been obtained for n coders in m languages. Our LD-metric L_i will then be defined as:

$$L_i = \frac{-2}{m(m-1)} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \frac{\left|M_{i,j} - M_{i,k}\right|}{C_{i,j,k}} \qquad \textbf{(2)}$$

The variable $C_{i,j,k}$ indicates the *critical interval* for statistical significance of the difference $|M_{i,j} - M_{i,k}|$. Hence, if all differences between each pair of tested languages are *just* statistically significant for coder i, then L_i will be equal to 1. Now we are left with the problem of calculating $C_{i,j,k}$. Critical intervals may be obtained by carrying out an appropriate statistical analysis, such as an analysis of variance. We use, more specifically, a post-hoc Duncan's range test.

The 95% confidence intervals associated with values of L_i only depend on the number of languages; the number of languages is equal to the number of independent observations on which the differences $|M_{i,j} - M_{i,k}|$ are based. Since each individual difference is (by definition) significant if the difference is larger then one, then the same 95% confidence is reached for L_i if the difference is larger than the difference given by equation (3).

$$\delta L_i = \sqrt{\frac{1}{m}} \qquad \textbf{(3)}$$

In our case, using four languages, the 95% confidence interval associated with all language dependency metrics is equal to 0.5. This does *not* imply that the metric can be measured with exactly equal precision for each coder. Differences in precision are also reflected in the value of the LD-metric L_i itself: when the critical intervals $C_{i,j,k}$ are greater, then lower values of the LD-metric are found.

The '-' sign in equation (2) makes sure that values of the LD-metric are always negative. This is only done to meet the requirement of the precision-weighted data combination procedure that higher performance is indicated by numerically higher values. The negative sign is necessary to convert the LD-metric from a 'penalty' to a 'performance' measure.

3.5 Talker recognizability test method

The TNO talker recognizability test method (Steeneken & Van Leeuwen, 1997) uses talkers and listeners who are employed at TNO Human Factors. The population of this institute consists of approx. 150 people, who interact in many different structured and unstructured ways on a daily basis. Generally, TNO-HF employees are capable of recognising most of their colleagues' voices.

For the talker recognizability experiment, specific selection criteria were imposed on talkers and listeners. Sixteen talkers (8 male, 8 female) were selected, each of whom worked at TNO-HF for over 7 years, in jobs which require them to frequently talk with people throughout the institute (such as the receptionist, a computer helpdesk employee and the controller). Additionally, 8 "unknown" talkers (4 male, 4 female) were recorded, not working at TNO-HF, but with global voices matching the 16 known talkers in terms of dialect, speaking rate, etc.

The listeners were also recruited from the TNO-HF. The listeners worked at the institute for a minimum of 2 years. Before the actual experiment, the listeners were confronted with names and photographs of the 16 known talkers. If a listener indicated beforehand that a certain talker was unknown to him or her, then in the analysis of the results this talker would be treated in the same way as the "unknown" talkers.
During the experiments each listener was seated in a sound proof booth. All speech stimuli were presented by headphones. Listeners were asked to identify the talker, based on speech stimuli of increasing length. All stimuli are taken from the following Dutch sentence:

"De zesenvijftigjarige man is van Turkse afkomst en woont al tientallen jaren in de gemeente"
(the 56 year old man is of Turkish origin and has lived in the town for decades)

Initially, only a short stimulus is presented:

"man"

If the talker is not immediately recognised, stimuli of increasing duration are presented:

"man"
"afkomst"
"de zesenvijftigjarige man"
"de zesenvijftigjarige man is van Turkse afkomst"
"De zesenvijftigjarige man is van Turkse afkomst en woont al tientallen jaren in de gemeente"

The subject can respond by clicking buttons on a computer screen. There are 16 buttons representing "known" talkers, showing the name and photograph of these talkers. There is also a button "unknown male" and a button "unknown female", and a button "Play longer fragment". By pressing this latter button, the subject requests to hear the next (longer) stimulus; after 5 presentations, this button is disabled and the listener is forced to make a choice. The scores for a correct response range from "1" to "5", representing stimuli of increasing duration. The penalty for incorrect responses was a score of 10. Hence, the better a subject is able to recognise a talker, the lower the score. By taking mean scores for a given vocoder (in a given condition) across all talkers, the talker recognizability when using that particular vocoder is quantified. By means of an ANalysis Of Variance (ANOVA), associated 95% confidence margins are calculated.

4 TEST RESULTS

4.1 CVC results

4.1.1 CVC results to be used for the selection procedure

In Table 2, mean equally-balanced CVC word scores and standard errors are given for all conditions. Each mean is based on the scores for 32 speaker-listener pairs.

Table 2 Mean equally-balanced CVC word scores (%) and standard errors, for coders 1–9 in conditions noise01–noise12 (N=32).

	Noise01		Noise02		Noise03		Noise04	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
US2400	66.5	2.2	27.7	1.8	48.5	1.8	59.6	1.8
CELP	66.7	1.6	30.5	1.5	46.5	1.9	61.9	1.5
FR2400	65.1	1.5	25.4	1.5	43.9	2.1	60.5	2.1
CVSD	67.2	1.9	28.9	1.9	43.4	2.2	58.1	2.2
TU2400	60.7	1.9	29.0	1.4	45.6	1.5	61.6	2.2
US1200	58.6	2.0	26.9	1.4	42.1	2.1	57.8	1.6
LPC-10	51.4	2.1	20.4	1.4	38.4	1.6	46.7	2.3
TU1200	57.7	1.7	19.0	1.6	35.9	2.0	50.4	1.6
FR1200	63.3	1.6	21.8	1.8	37.9	2.1	55.5	2.1
	Noise05		**Noise06**		**Noise07**		**Noise08**	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
US2400	59.2	2.5	39.2	2.1	37.9	1.8	50.7	2.0
CELP	62.9	2.3	30.3	1.3	36.7	1.2	40.8	2.0
FR2400	55.7	2.1	25.4	2.0	30.2	1.9	44.3	2.0
CVSD	58.0	2.5	32.8	2.0	36.2	1.8	44.1	1.9
TU2400	56.6	2.3	33.2	1.3	38.8	1.5	53.2	1.8
US1200	58.6	2.2	37.8	1.9	37.0	1.5	45.2	2.0
LPC-10	47.4	2.0	18.6	1.6	18.9	1.0	29.0	1.7
TU1200	51.5	1.8	33.5	1.5	29.2	1.5	39.7	1.8
FR1200	56.3	1.6	26.2	1.8	26.7	1.2	39.2	1.6
	Noise09		**Noise10**		**Noise11**		**Noise12**	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
US2400	64.9	1.9	54.2	2.0	60.8	1.9	42.0	2.9
CELP	61.5	2.3	56.2	1.5	51.8	1.8	44.4	2.3
FR2400	64.8	1.7	49.8	1.7	56.4	1.6	43.0	1.9
CVSD	62.2	2.0	42.4	2.4	56.1	2.4	56.4	2.0
TU2400	65.1	1.8	56.5	1.8	51.2	1.9	40.5	1.7
US1200	58.9	2.1	53.6	1.9	46.0	2.0	41.8	1.8
LPC-10	52.2	2.1	34.4	1.7	43.9	1.8	27.2	1.9
TU1200	55.3	2.0	47.7	2.1	47.9	1.6	34.1	1.9
FR1200	57.9	2.0	46.6	1.9	51.1	2.1	33.8	1.9

The CVC word scores for the different coders form a consistent pattern across the various acoustic noise conditions.

4.1.2 Additional CVC results (informative)

Speaker gender is quite often an important variable for coder performance. In Table 3, mean equally-balanced CVC word scores and standard errors are given for male speakers and female speakers separately.

Table 3 Mean equally-balanced CVC word scores (%) for male and female speakers separately, for coders 1–9 in conditions noise01–noise12. The average standard error corresponding to the mean values in this table is 2.3 (N=16).

	Noise01		Noise02		Noise03		Noise04	
	Male	Female	Male	Female	Male	Female	Male	Female
US2400	72.8	60.2	32.5	22.9	54.5	42.4	63.8	55.3
CELP	71.3	62.0	32.0	29.1	50.3	42.8	64.8	59.1
FR2400	66.9	63.4	29.3	21.6	50.6	37.2	66.5	54.5
CVSD	72.4	61.9	31.0	26.8	50.9	36.0	65.9	50.4
TU2400	62.5	58.9	32.4	25.7	47.7	43.6	70.1	53.2
US1200	67.4	49.9	32.2	21.6	45.8	38.4	60.4	55.2
LPC-10	56.2	46.5	21.6	19.3	40.2	36.5	53.8	39.7
TU1200	58.7	56.7	21.6	16.4	36.6	35.2	53.4	47.4
FR1200	67.3	59.3	27.3	16.2	41.7	34.2	60.1	50.9
	Noise05		**Noise06**		**Noise07**		**Noise08**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
US2400	68.3	50.1	41.8	36.5	41.1	34.7	56.6	44.7
CELP	69.1	56.6	32.4	28.3	38.7	34.7	47.9	33.7
FR2400	61.6	49.7	33.6	17.3	36.6	23.8	50.6	38.0
CVSD	65.2	50.7	39.1	26.5	40.8	31.6	50.1	38.0
TU2400	61.9	51.3	37.4	28.9	40.6	37.0	59.7	46.7
US1200	62.3	54.9	40.7	34.9	39.3	34.7	51.0	39.5
LPC-10	54.9	39.8	18.5	18.8	19.6	18.3	32.5	25.5
TU1200	54.5	48.5	37.0	29.9	29.5	28.8	46.0	33.5
FR1200	54.0	58.5	30.1	22.2	29.5	23.8	44.1	34.2
	Noise09		**Noise10**		**Noise11**		**Noise12**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
US2400	70.7	59.2	58.3	50.0	67.0	54.5	52.9	31.1
CELP	68.7	54.3	57.2	55.3	57.5	46.2	51.1	37.8
FR2400	68.1	61.4	53.9	45.6	60.2	52.7	49.1	36.9
CVSD	66.3	58.1	50.0	34.8	63.1	49.0	61.3	51.5
TU2400	68.4	61.9	58.7	54.3	49.0	53.4	47.4	33.6
US1200	65.7	52.1	59.8	47.3	54.3	37.7	46.8	36.8
LPC-10	57.8	46.5	31.1	37.8	47.2	40.7	33.2	21.1
TU1200	58.0	52.6	49.0	46.3	49.0	46.8	37.4	30.9
FR1200	64.2	51.6	47.1	46.1	58.3	43.9	41.3	26.2

From the CVC responses, full confusion matrices may be calculated. In Annex A such confusion matrices are given for all 9 coders, pooled across conditions. Separate matrices are given for initial consonants, vowels and final consonants.

4.2 MOS results

4.2.1 MOS results to be used for the selection procedure

In Table 4, mean MOS scores and standard errors are given for all test conditions. Each mean is based on the scores for 152 talker-listener pairs (8 talkers, 19 listeners).

Table 4 Mean MOS scores and standard errors, for coders 1–9 in conditions noise01–noise10 (N=152).

	Noise01		Noise02		Noise03		Noise04		Noise05	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
US2400	4.22	0.063	1.96	0.079	2.84	0.070	3.09	0.101	3.38	0.072
CELP	3.83	0.067	2.03	0.073	2.73	0.060	3.32	0.066	3.34	0.070
FR2400	3.72	0.064	1.55	0.058	2.39	0.086	3.20	0.079	3.01	0.066
CVSD	3.24	0.056	2.42	0.080	2.91	0.069	3.24	0.058	3.13	0.069
TU2400	3.11	0.065	1.51	0.057	2.23	0.079	2.82	0.064	2.82	0.068
US1200	3.57	0.060	1.84	0.068	2.80	0.066	3.26	0.077	2.99	0.073
LPC-10	2.48	0.059	1.30	0.048	2.00	0.071	2.22	0.055	2.25	0.063
TU1200	2.76	0.072	1.31	0.050	2.11	0.069	2.37	0.058	2.44	0.065
FR1200	3.47	0.062	1.47	0.065	2.21	0.069	2.56	0.059	2.79	0.071
	Noise06		**Noise07**		**Noise08**		**Noise09**		**Noise10**	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
US2400	2.13	0.066	3.63	0.057	2.91	0.071	3.04	0.062	2.74	0.057
CELP	2.06	0.075	3.30	0.063	2.68	0.070	2.34	0.070	2.82	0.055
FR2400	1.68	0.067	3.41	0.075	2.63	0.072	2.93	0.066	2.79	0.060
CVSD	2.41	0.086	3.17	0.062	2.95	0.077	2.73	0.079	2.86	0.064
TU2400	1.89	0.062	3.05	0.070	2.46	0.070	2.39	0.075	2.36	0.058
US1200	2.01	0.075	3.35	0.057	2.76	0.066	2.56	0.068	2.72	0.060
LPC-10	1.26	0.049	2.15	0.058	1.31	0.049	1.88	0.055	2.01	0.058
TU1200	1.55	0.052	2.56	0.066	2.22	0.061	2.10	0.057	2.03	0.063
FR1200	1.46	0.054	2.97	0.069	2.40	0.067	2.47	0.075	2.48	0.056

4.2.2 Additional MOS results (informative)

The responses given in the MNRU calibration conditions can be used to verify the validity of the test. The MNRU responses should ideally cover the full MOS scale, and a monotonous relation between MNRU SNR (in dBQ) and the MOS score should be observed. The relation between MNRU SNR and mean MOS score (152 speaker-listener pairs per measuring point) is given in Figure 1.

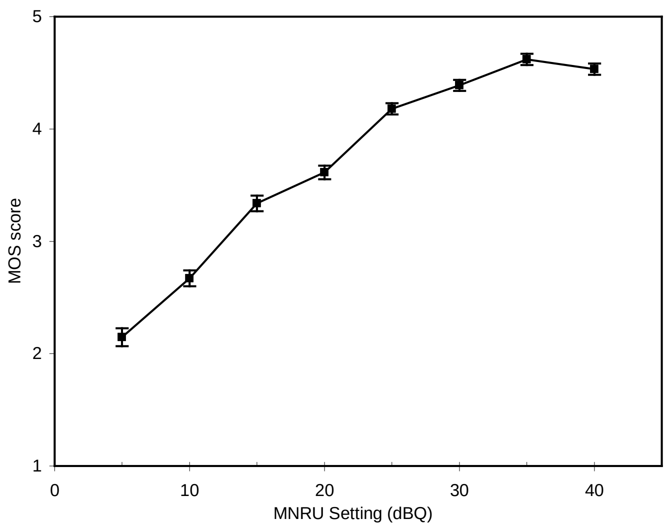

Fig. 1 Relation between MNRU SNR (dBQ) and mean MOS score. The error bars indicate the standard error (152 speaker-listener pairs per measuring point).

There is a noticeable ‘floor-effect’ in the MNRU-curve: since there are relatively many heavily distorted speech samples among the test conditions, the lowest dBQ-condition (5 dBQ) does not completely approach a MOS score of 1. This is commonly observed when applying MOS tests to narrow band voice coders. There appears to be a slight decrease in MOS from 35 to 40 dBQ, but this difference is not statistically significant.

4.3 Whispered speech results

4.3.1 Intelligibility of whispered speech (results to be used for the selection procedure)

SRT tests with whispered speech by 4 talkers (2 male, 2 female) were conducted with 12 listeners. In addition to speech processed through each of the 9 candidate vocoders, a condition with unprocessed speech (bandwidth limited 0–4000 Hz) was also tested.
In Table 5, normalised inverse SRT scores are presented, along with the corresponding standard errors. The scores represent robustness of whispered speech to additive noise (*after* vocoder processing), relative to the mean of all tested listener-talker-coder combinations. The SRT results are presented in this somewhat unconventional form to be directly applicable in the precision-weighted data combination method.

Table 5 Mean inverse SRT scores and standard errors, for coders 1–9 (N=12). The scores are normalised for between-talker and between-listener differences.

	Mean	S.E.
Unprocessed speech	7.25	0.52
US2400	3.35	0.65
CELP	4.35	0.39
FR2400	–1.17	0.79
CVSD	3.32	0.60
TU2400	–0.12	0.64
US1200	–4.99	0.66
LPC-10	–8.55	0.74
TU1200	–3.11	1.06
FR1200	–0.34	0.73

Most of the differences between any of the coders are statistically significant. The performance of all coders differs significantly from unprocessed speech. Due to the relatively small number of statistically independent observations (N=12), it is necessary that the appropriate Student-t value is taken from the Student-t distribution for the calculation of 95% confidence intervals.

4.3.2 Additional information regarding the intelligibility of whispered speech (informative)

A disadvantage of the form in which the SRT results are presented in Table 5, is that it is difficult to compare these results to results from future and past SRT tests concerning vocoders. Although the differences between the vocoders are fully preserved by the normalisation operations, the absolute meaning of the SRT-values is lost.
For this reason, the ‘raw’ SRT scores are also reported in Table 6.

Table 6 Mean 'raw' SRT scores and standard errors, for coders 1–9 (N=12). Each score gives the average speech-to-noise ratio corresponding to 50% intelligibility of whispered sentences for that particular coder.

	Mean SRT (dB SNR)	S.E.
Unprocessed speech	3.97	0.52
US2400	8.67	1.31
CELP	7.03	0.73
FR2400	12.77	1.49
CVSD	8.13	0.88
TU2400	11.33	0.63
US1200	17.00	1.12
LPC-10	19.93	1.07
TU1200	14.70	1.24
FR1200	11.80	1.04

It is interesting to compare the SRT score for unprocessed whispered speech to the same value for normal speech. For normal speech, this value is 0.40 dB (standard error 0.47 dB; result taken from the Dutch part of the language dependency experiment, which is presented in the following section). Hence, the baseline effect of whispering on speech intelligibility in terms of speech-to-noise ratios is 3.5 dB. For the coders, this difference can be much greater; up to approx. 15 dB for LPC-10. This indicates that some vocoders suffer a disproportional loss of speech intelligibility when the talker is whispering.

4.4 Language dependency results

4.4.1 Language dependency results to be used for the selection procedure

Using the method described in section 3.4, individual SRT results obtained in four languages (English, French, German and Dutch) were used to calculate the TNO language dependency metric.

Table 7 Mean inverse SRT scores and standard errors, for coders 1–9 (N=12). The scores are normalised for between-talker and between-listener differences.

	LD-metric	95% confidence interval
US2400	–0.226	0.5
CELP	–0.516	0.5
FR2400	–0.248	0.5
CVSD	–0.256	0.5
TU2400	–0.971	0.5
US1200	–0.172	0.5
LPC-10	–0.469	0.5
TU1200	–0.383	0.5
FR1200	–0.396	0.5

The closer the scores in Table 7 are to zero, the less language dependent the corresponding vocoder is. The metric is (artificially) given a negative sign for all coders, in order to have better performance lead to higher scores, as required by the precision-weighted data combination procedure.

4.4.2 Additional results from the language dependency test (informative)

All values of the LD-metric in Table 7 were calculated from normalised SRT scores using equation (2) from section 3.4. The normalisation of the 'raw' SRT scores was applied to remove small systematic differences in SRT scores due to differences in test implementations across languages, differences in talkers' clarity of speaking, and differences in individual subjects' aptness at the SRT task.

Table 8 Mean 'raw' SRT-scores and standard errors (N=10), for each coder-language combination.

	English		French		German		Dutch	
	mean	**S.E.**	**mean**	**S.E.**	**mean**	**S.E.**	**mean**	**S.E.**
Unprocessed speech	0.68	0.38	1.04	0.78	0.25	0.35	0.40	0.47
US2400	3.24	0.34	3.32	0.81	2.32	0.69	2.52	0.55
CELP	2.80	0.54	4.16	0.65	2.12	0.33	2.52	0.60
FR2400	3.16	0.63	4.00	0.63	3.12	0.35	2.88	0.60
CVSD	2.48	0.37	2.56	0.44	1.64	0.49	1.68	0.28
TU2400	1.40	0.37	2.32	0.44	2.88	0.63	2.36	0.67
US1200	4.24	0.67	4.64	0.85	3.44	0.39	3.88	0.58
LPC-10	4.72	0.57	3.96	0.34	3.76	0.39	4.12	0.51
TU1200	3.04	0.64	4.27	0.56	3.24	0.54	3.16	0.82
FR1200	3.52	0.73	4.81	0.88	3.32	0.42	3.68	0.54

In Table 8, mean 'raw' SRT scores for all coder-language combinations are given. Differences in mean SRT scores for 'unprocessed speech' across languages indicates systematic differences in the SRT implementation across languages. However, none of these differences in Table 8 are statistically significant.

Table 8 also shows that there is a good general agreement of the SRT test across languages. This indicates that language dependency is a minor issue with most of the vocoders; generally, the same SRT results are found irrespective of the test language. This is also illustrated by Figure 2.

The total explained variance in the relations between SRT results across languages is relatively high. In the relation between Dutch and German SRT results the explained variance is even 91%, which indicates that the vocoder performance is very similar for these two languages.
Language dependency should become manifest in Figure 2 in the form of data points that appear at some distance from the regression line. In the panel with the correlation between German and English, such a point is indicated as a solid instead of an open square, corresponding to TU2400. According to Table 7, the most language dependent vocoder is TU2400. The solid square in Figure 2 shows part of the reason for this.

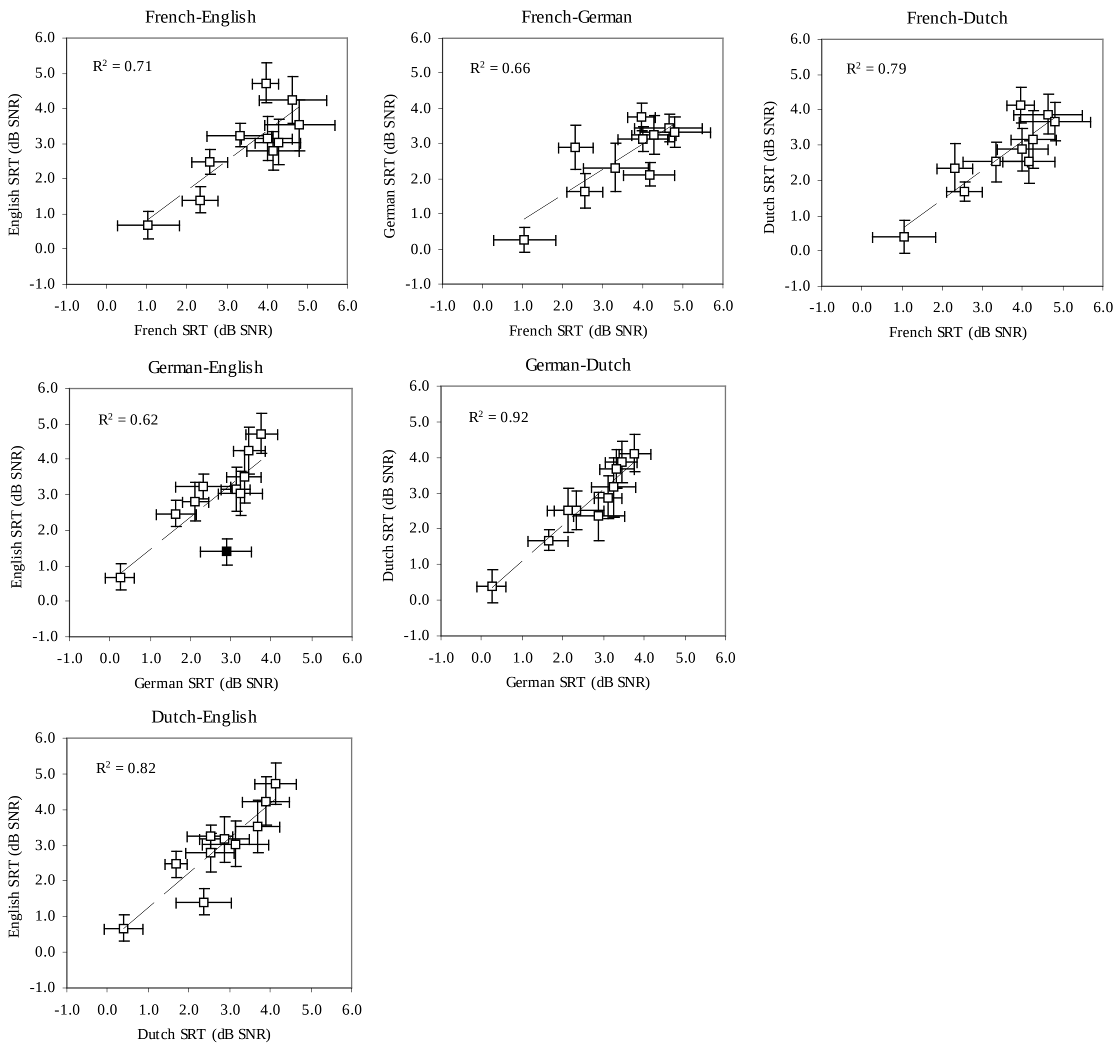

Fig. 2 Comparison of mean SRT scores between all possible pairs of languages, for all 9 coders. The error bars indicate the standard errors of the raw SRT scores (N=10). The squared correlation coefficient (explained variance) between languages is also indicated.

4.5 Talker recognizability results to be used for the selection procedure

The talker recognizability test results in a penalty score for each vocoder; the score is higher if the corresponding vocoder makes it more difficult to recognise talkers. Because of the requirements of the precision-weighted data combination procedure, the results given in Table 9 represent the *negative* of the original penalty-scores.

Table 9 Mean talker recognizability scores and associated 95% confidence intervals (Duncan's test), for each coder in two conditions.

	Noise01		Noise02	
	mean	95% C.I.	mean	95% C.I.
US2400	–4.33	0.350	–4.89	0.432
CELP	–4.10	0.327	–4.74	0.403
FR2400	–4.51	0.367	–5.38	0.453
CVSD	–4.16	0.335	–4.70	0.393
TU2400	–4.35	0.360	–4.86	0.420
US1200	–4.77	0.373	–5.36	0.448
LPC-10	–5.62	0.384	–7.54	0.461
TU1200	–4.89	0.378	–5.19	0.441
FR1200	–5.16	0.382	–5.96	0.458

4.6 Relation between performance measures (informative)

Speech quality and speech intelligibility are fairly correlated measures, although the degree of correspondence between these measures depends on the vocoder characteristics and the noise environment. After phase 1 of the selection procedure, the correlation between CVC word score and MOS scores was reported for all vocoders, in noise conditions Noise01–Noise04 (Van Wijngaarden *et al.*, 2001c). Figure 3 shows this same correlation, now based on phase 2 results for noise01–noise04.

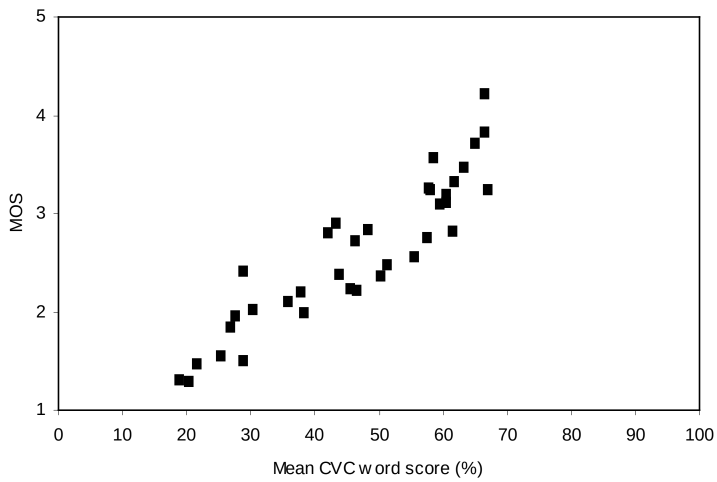

Fig. 3 Correlation between mean CVC word scores and MOS-scores. Each data point represents means scores over all talkers and listeners for a coder-by-a-condition (noise01–noise04; R=0.92).

The results given in Figure 3 are similar to those of phase 1; there is a good correlation between MOS and CVC scores, confirming the expectations regarding the behaviour of the test methods. In Figure 4, the same correlation is now given across *all* conditions that the CVC and MOS tests had in common.

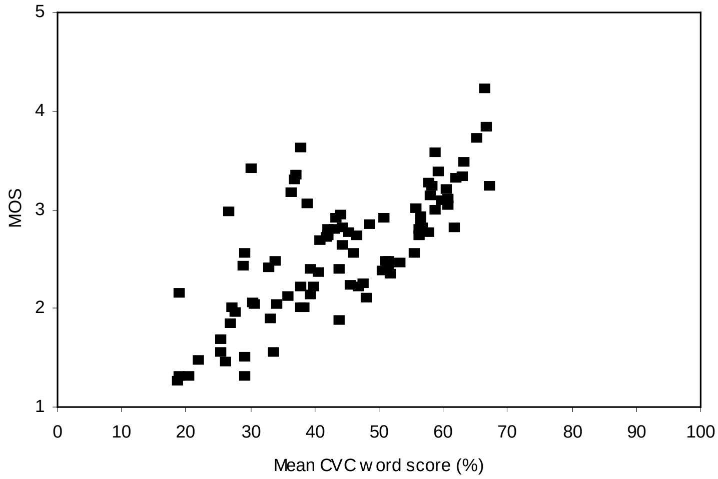

Fig. 4 Correlation between mean CVC-word scores and MOS-scores. Each data point represents means scores over all talkers and listeners for a coder-by-a-condition (all common conditions between MOS and CVC; R=0.73).

The correlation coefficient corresponding to Figure 4 is smaller than the correlation coefficient corresponding to Figure 3. This indicates that there are dissimilarities between the CVC and MOS scores for some of the tested conditions.

Another good correlation is expected between CVC intelligibility results and SRT whispered speech results. This correlation is given in Figure 5.

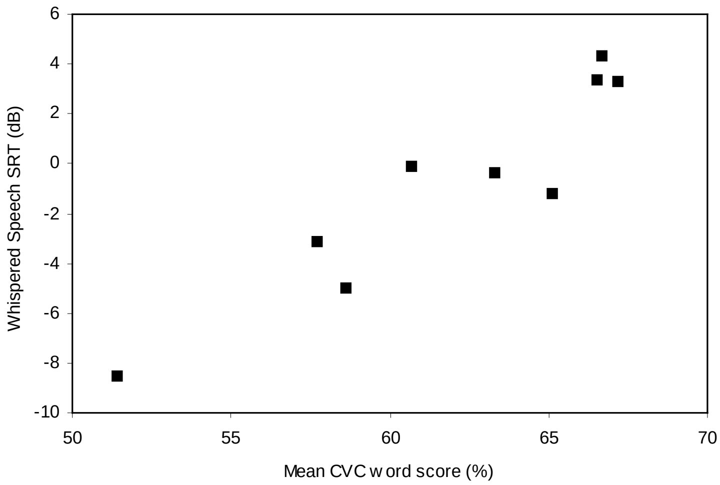

Fig. 5 Correlation between mean CVC-word scores and whispered speech SRT-scores. All test results correspond to the condition without adding acoustic noise before vocoder processing (R=0.94).

5 CONCLUSIONS

Five different performance characteristics of vocoders were measured, in a number of conditions. The overall set of test scores appears as a coherent body of data; routine ‘health checks’ of the data uncovered no reasons to question the validity of the outcome.

The highest average intelligibility across all conditions is offered by US2400, followed by CELP and TU2400. According to MOS results, US2400 also offers the highest overall speech quality. CELP and TU2400 were found to be significantly language dependent.

REFERENCES

CCITT (1989). *Recommendation P.810, Modulated Noise ReferenceUnit (MNRU).* Vol.of Blue Book, pp. 198-203, Geneva: ITU.

Plomp, R. & Mimpen, A.M. (1979), Improving the reliability of testing the speech reception threshold for sentences. *Audiology, 18,* 43-52.

Steeneken, H.J.M., Geurtsen, F.W.M & Agterhuis, E. (1990). *Speech Data-Base for Intelligibility and Speech Quality Measurements* (Report IZF 1990 A-23). Soesterberg, NL: Institute for Perception TNO.

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

Steeneken, H.J.M. & Leeuwen, D.A. van (1997). Speaker Recognition by Humans and Machines. *Proceedings Eurospeech97,* Rhodes, Greece, 2319-2322.

Tardelli et al. (2000). Test and selection plan 2400bps/1200bps Digital Voice Coder. Version 1.012, *NATO Ad-Hoc Working Group AC/322(SC/6-AHWG/3),* May 2000.

Velden, J.G. van (1991). *Speech level meter, version 2, SAM_SLM* (Esprit-SAM, document SAM-TNO-043). Soesterberg, The Netherlands: Institute for Perception TNO.

Wijngaarden, S.J. van & Steeneken, H.J.M. (1998). The evaluation of speech transmission quality. *Proceedings AC/243 (Panel 8) TP/17,* NATO.

Wijngaarden, S.J. van & Steeneken, H.J.M. (2001a). A Proposed Method for Measuring Language Dependency of Narrow Band Voice Coders. *Proceedings Eurospeech 2001,* Ålborg, Denmark, pp. 2495-2498.

Wijngaarden, S.J. van, Steeneken, H.J.M. & Houtgast, T. (2001b). Methods and models for quantitative assessment of the intelligibility of nonnative speech. *Proceedings of the NATO-RTO workshop on Multi-Lingual Speech Processing,* Ålborg, Denmark.

Wijngaarden, S.J. van, Buuren, R.A. van & Steeneken, H.J.M. (2001c). *NATO-NBVC 1200/2400 bps coder selectionresults of TNO tests in phase 1* (Report TM-01-C005). Soesterberg, The Netherlands: TNO Human Factors.

Soesterberg, 29 October 2001

Ir. S.J. van Wijngaarden
(First author, Project leader)

ANNEX A Confusion matrices

The patterns of phone confusions that occur when using vocoders, will be a result of at least three different factors. First of all, there is a natural tendency of subjects to discriminate between some phoneme pairs better than between others. Secondly, the particular noise condition has an influence. Thirdly, and most importantly, there is the influence of the coder itself, which may represent some phonemes better than others. It is this particular effect which makes studying phoneme confusions worthwhile. The confusion matrices given in this annex are compiled from *all* responses for a given coder; they contain pooled results over all noise conditions. For each coder, three different matrices are given, corresponding to confusions of the initial consonant, the vowel, and the final consonant in each CVC word.
Table A.0.1 is an example of a confusion matrix, to explain the meaning of the values found in the tables of this annex.

Table A.0.1 Example of a confusion matrix.

	p	t	k	??	Tot	%
p	288	14	25	5	384	75.0
t	19	312	33	7	384	81.3
k	19	13	330	3	384	85.9
conf	38	27	58	15		

The example table shows confusion between only three consonants: /p/, /t/, an /k/. Following regular conventions, each row contains all responses to a certain stimulus. In other words: the stimuli form the vertical axis of the matrix, the response categories the horizontal axis. Since the CVC-test is an open-response test, a fraction of the responses can not be classified according to the response categories. These are responses that do not meet the CVC word structure, or indicate phonemes that are not included in the CVC test, or perhaps no response at all. The column indicated by '??' lists all these irregular responses to each stimulus.
The column indicated by 'Tot' gives the total number of presentations for each stimulus; the column indicated by '%' gives the percentage of correct responses for each phone (stimulus). The bottom row ('conf') gives the number of confusions within each response category.
Some confusions are very common; examples are the /s/–/z/ and /f/–/v/ confusions. Similar phones are listed in the tables close to each other; for instance, the plosives form the top-left part of the confusion matrix of the tables for consonants. This implicates that confusions appearing close to the diagonal of the matrix should be expected to occur more frequently than confusions appearing away from the diagonal.

Table A.1.1 Confusion matrix of initial consonants for US2400.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	832	111	88	31	5	21		17		22				1	8	1	14	1	1152	72.2
t	72	916	99	3	21	9	1	4	6	8		1		2	4		6		1152	79.5
k	27	27	1045		1	5		4		32				2		2	7		1152	90.7
b	86	5	9	607	84	4		6		2	3	4	4	11	300	15	11	1	1152	52.7
d	18	87	13	38	836	1		4	3	2	1	2	2	14	98	29	8		1156	72.3
f	43	14	10	3	1	599	24	338	3	96		1			5	1	14		1152	52.0
s						26	997	17	95	15				1			1		1152	86.5
v	34	20	15	11	7	271	5	536	22	77	2		3	18	92	17	22		1152	46.5
z		1	1		8	4	180	27	831	3		3	2	9	19	56	8		1152	72.1
χ	14	2	6	1		36	1	29	1	1034		1			2	7	18		1152	89.8
m				5	1						615	368	33	10	62	33	24	1	1152	53.4
n			1	2	2						102	916	20	9	21	69	5	1	1148	79.8
l			3	2	2				4	1	18	154	698	50	48	150	21	1	1152	60.6
r	4	1	1	5	4	2		6	6	27	4	6	27	936	52	17	53	1	1152	81.3
ʋ	16		1	90	16	1		5	6		14	12	32	48	834	38	39		1152	72.4
j	1		6	1	7			1	7		3	9	11	5	28	1055	17	1	1152	91.6
h	21	1	10	16	5	3	1	13	2	45	5	11	7	47	22	16	926	1	1152	80.4
conf	336	269	263	208	164	383	212	471	155	330	152	572	141	227	761	451	268	8		

Table A.1.2 Confusion matrix of vowels for US2400.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1081	22	26		3	3	2					1	1	7	2		1148	94.2
ɑu	16	868	252		2		1			11				5		1	1155	75.2
ɑ	28	86	2116				12			8	3	47		3		1	2303	91.9
eː	1			980	1	5	7	18	131	1		1			7		1152	85.1
ɛi	1			31	##	1	40	1	2					8	1	2	1150	92.6
øː		1		36	1	876	5	1	25	7	13	10	12	10	152	3	1149	76.2
ɛ	2	8	6	36	67	3	1913	9	187		5	2	1	21	42	2	2302	83.1
iː				23		1	2	##	58		25		35		5		1152	87.1
ɪ				74		3	30	138	869		8	1	9		20		1152	75.4
oː		12	5		1	3	1			826	29	273			1	1	1151	71.8
uː			1	5				26	7	16	986	69	29		12	1	1151	85.7
ɔ			13				1	1	11	75	72	967	3		8	1	1151	84.0
yː				3		9		119	13		140	3	828		33		1148	72.1
œy	7	5	7	1	52	11	27		2		1		1	1041	1		1156	90.1
ʉ				5		70	12	19	71	2	41	29	68		835		1152	72.5
conf	55	134	310	214	127	109	140	332	507	120	337	436	159	54	284	12		

Table A.1.3 Confusion matrix of final consonants for US2400.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	781	154	162	20	6	16	1		2		7	3	1152	67.8
t	107	2028	91	36	10	15	1	3	2	2	6	3	2304	88.0
k	40	68	988	9	2	26	2				15	2	1152	85.8
f	30	42	21	839	51	152	2			8	5	2	1152	72.8
s		11	4	53	2205	25					4	2	2304	95.7
χ	12	13	14	85	24	985		1	1	8	9		1152	85.5
m	18		3	2	1	3	1344	172	696	36	21	8	2304	58.3
ŋ	5	6	5	1			191	711	203	22	6	2	1152	61.7
n	11	2		1	1	3	674	223	1283	75	22	9	2304	55.7
l	5	1	3		2		25	10	36	2052	58	112	2304	89.1
r	9	8	15	21	3	36	5	6	10	36	2154	1	2304	93.5
conf	237	305	318	228	100	276	901	415	950	187	153	144		

Table A.2.1 Confusion matrix of initial consonants for CELP.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	747	136	142	15	8	25	1	17		18	2	1		5	8	4	22	1	1152	64.8
t	106	838	109	3	35	26	1	15		6		1		3	3		6		1152	72.7
k	72	51	959	3	1	6		2	1	42				1	1	1	12		1152	83.2
b	36	7	28	697	153	2		5		1	5	3	4	13	183	11	4		1152	60.5
d	7	54	11	76	878	6		4	2	3		6	5	7	50	35	8		1152	76.2
f	107	40	70	4		517	11	269	2	107				2	3		8		1140	45.4
s	4	5	3			26	999	23	72	19					1				1152	86.7
v	47	24	25	21	19	251	11	474	33	73	1	1	4	28	122	13	17		1164	40.7
z		1	1	1	14		181	11	821	10			3	3	26	73	5	2	1152	71.3
χ	18	7	43	2	1	20	10	10		1023				1	1	1	15		1152	88.8
m			1	7	2						627	362	55	9	40	33	16		1152	54.4
n				2	3						103	876	72	10	20	64	1	1	1152	76.0
l	2	2	1	3	10				2		15	89	813	46	32	127	10		1152	70.6
r	6	4	1	9	5			11	8	14	3	6	11	987	45	20	21	1	1152	85.7
ʋ	7	1	1	69	24			3	6		29	18	36	28	844	57	29		1152	73.3
j			6	2	7			1	11		2	15	23	6	39	1030	10		1152	89.4
h	14	2	14	9	8	6		14	4	63	5	8	8	45	38	34	880		1152	76.4
conf	426	334	456	226	290	368	215	385	141	356	165	510	221	207	612	473	184	5		

Table A.2.2 Confusion matrix of vowels for CELP.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1056	24	37	1	1		1	3			1			13		3	1137	92.9
ɑu	29	832	262		2		2			18		1		8	1	1	1155	72.0
ɑ	27	48	2155			1	8	1		14	7	33		6	1	3	2301	93.7
eː				992	2	18	10	21	104	2	1	6	1		3		1160	85.5
ɛi	3	1	1	29	1042		53	1	3					17	2		1152	90.5
øː		1		28	1	891	8	1	8	12	7	9	11	14	161		1152	77.3
ɛ	3	5	13	26	54	2	1999	8	116		3	5	4	33	32	1	2303	86.8
iː				45		1	2	961	67		15	1	55		5		1152	83.4
ɪ				50	1	2	47	105	895		5	2	10		35		1152	77.7
oː		14	12	5	2	7	1			879	18	202		4	4		1148	76.6
uː		1	1	5		11		32	12	34	941	28	53		38		1156	81.4
ɔ		1	50			3	7	4	3	69	32	971	1		11		1152	84.3
yː				1		17		97	23		132	2	801	8	51		1132	70.8
œy	24	11	17	2	30	15	17			4				1039	4	1	1163	89.3
ʉ				6		82	28	7	105	4	20	24	63		820	1	1159	70.8
conf	86	106	393	198	93	159	184	280	441	157	241	313	198	103	348	10		

Table A.2.3 Confusion matrix of final consonants for CELP.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	737	153	220	15	1	11	2	1	1	3	7	1	1152	64.0
t	161	1891	152	25	33	26	1	1	4		6	4	2304	82.1
k	81	67	965	5	2	19	2	1	4	2	4		1152	83.8
f	85	99	89	635	50	164	6		3	5	15	1	1152	55.1
s	3	16	9	28	2229	15				2	1	1	2304	96.7
χ	18	29	34	40	23	987		1	2	5	11	2	1152	85.7
m	5	3	4			4	1427	145	653	44	11	8	2304	61.9
ŋ		2	4	1		2	171	723	212	27	8	2	1152	62.8
n	3	4	3			1	737	222	1273	42	17	2	2304	55.3
l	2	5	6	2		1	17	6	22	2057	91	95	2304	89.3
r	20	14	45	12	1	22	5	4	15	29	2135	2	2304	92.7
conf	378	392	566	128	110	265	941	381	916	159	171	118		

Table A.3.1 Confusion matrix of initial consonants for FR2400.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	υ	j	h	??	Tot	%
p	851	85	104	17	2	19		13		8		1	1		12	5	33	1	1152	73.9
t	102	896	76	8	19	7	2	7	4	5	1	2	1	1	3	9	8	1	1152	77.8
k	22	18	1060	1	2	1	3	2		11	1		3		2	3	23		1152	92.0
b	97	8	16	497	73	4		7	1	8	7	14	12	11	364	5	26	2	1152	43.1
d	27	70	13	61	760	5	1	2	10	9		8	9	11	95	54	24	1	1160	65.5
f	98	22	29	3	3	535	37	283	6	104			1	2	3	1	24	1	1152	46.4
s	7	1		1		23	973	21	94	21			1		2		3	1	1148	84.8
v	78	11	27	28	9	255	17	402	23	78	4	5	3	27	98	15	72		1152	34.9
z	3	6	2	4	12	8	180	26	782	4	1	3	10	13	18	71	11	2	1156	67.6
χ	26	4	26			37	12	14	2	992	1			6	1	1	29	1	1152	86.1
m				4	1			1			603	341	58	15	50	64	12	3	1152	52.3
n	1		2	4	8					1	92	883	45	8	19	62	18	1	1144	77.2
l	1	1	2	5		1	1		4		30	131	752	57	31	102	32	2	1152	65.3
r	6	1	12	8	5	1	1	7	4	21	5	15	22	882	57	36	68	1	1152	76.6
υ	17	1	4	53	18	1	1	14	2	5	25	14	37	49	811	39	60	1	1152	70.4
j	1	1	6	1	17	1	2		6	1	6	17	15	13	26	1017	22		1152	88.3
h	16	4	23	9	5	6	3	7	3	47	16	21	8	38	31	57	856	2	1152	74.3
conf	502	233	342	207	174	369	260	404	159	323	189	572	226	251	812	524	465	20		

Table A.3.2 Confusion matrix of vowels for FR2400.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1082	11	29		4		2					1		22		1	1151	94.0
ɑu	38	840	249		1	1				9				21	1		1160	72.4
ɑ	26	73	2130			1	15	3	2	9	3	30		8	1	3	2301	92.6
eː				948	38	3	14	39	99	2	2	1	1		3	2	1150	82.4
ɛi				24	1078	5	18	3	7			2		12		3	1149	93.8
øː	1			23	6	879	14	3	14	10	6	5	15	25	149	2	1150	76.4
ɛ	6	3	10	55	89	17	1897	26	123		3		12	23	40		2304	82.3
iː				52	1		1	956	75		19	2	40		6		1152	83.0
ɪ				107		5	86	117	800		2	3	6		24	2	1150	69.6
oː		30	24	2		4				797	31	254	1	1	4		1148	69.4
uː	1		3	4		3		32	12	33	913	75	53	1	17	1	1147	79.6
ɔ	1	2	54		1		7	1	6	80	52	936	2		7	3	1149	81.5
yː			1	3	1	35	2	103	32	1	87	4	787	1	93	2	1150	68.4
œy	24	16	8		32	20	20	1		3		3	4	1019	1	1	1151	88.5
ʉ	2		1	10	2	89	27	28	63	4	15	19	54	2	836		1152	72.6
conf	99	135	379	280	175	183	206	356	433	151	220	399	188	116	346	20		

Table A.3.3 Confusion matrix of final consonants for FR2400.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	702	170	236	10		17	1		2	7	4	3	1152	60.9
t	141	1983	109	18	11	22	2		3	4	5	6	2304	86.1
k	59	53	1005	4		23				4	1	3	1152	87.2
f	60	79	75	661	118	140				7	8	4	1152	57.4
s	2	23	5	38	2209	16					9	2	2304	95.9
χ	31	11	43	41	50	962			1	3	12	2	1156	83.2
m	10	7	4	4	1	7	1289	191	692	55	34	10	2304	55.9
ŋ	4	3	6			4	184	679	227	19	20	6	1152	58.9
n	4	2	5	3		4	730	199	1229	68	51	9	2304	53.3
l	3	9	5			5	37	7	42	2030	63	103	2304	88.1
r	23	6	27	4		15	6	9	17	41	2148	4	2300	93.4
conf	337	363	515	122	180	253	960	406	984	208	207	152		

Table A.4.1 Confusion matrix of initial consonants for CVSD.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	810	139	111	11	13	23		20		16		1			1	1	4	2	1152	70.3
t	94	885	121	1	29	5	7	3	2	4					1				1152	76.8
k	41	36	1058		3	5		2		13			1				1		1160	91.2
b	49	12	2	764	130	4		3	1		12	2	8	1	155	2	7		1152	66.3
d	7	44	5	65	948	3		3	15	2	1	1	3	3	28	18	5	1	1152	82.3
f	181	65	83	4	3	446	38	216	3	82				1	8		9	1	1140	39.1
s	29	81	56	1	1	68	698	48	90	58				3	2		3	2	1140	61.2
v	64	25	36	55	24	159	21	385	59	60		1	7	32	162	5	68	1	1164	33.1
z	1	4	3	4	58	11	100	26	673	8	1	4	4	15	90	109	52	1	1164	57.8
χ	27	7	37	4		13	5	5		1026				3			17		1144	89.7
m	1			3							697	309	53	1	18	51	15		1148	60.7
n	1	1			2						119	880	53	5	4	75	16		1156	76.1
l				2	1						13	90	979	12	15	35	5		1152	85.0
r	2	2	1	9	5	5	1		2	2	1	2	4	1042	27	6	41		1152	90.5
ʋ	5		3	69	25	2		5	10		12	11	37	49	827	56	39	2	1152	71.8
j	1	2	1	2	9	1	1		11	1		18	9	2	50	1021	22	1	1152	88.6
h	17	7	14	11	3	3		3	6	25	5	12	3	16	25	37	965		1152	83.8
conf	520	425	473	241	306	302	173	334	199	271	164	451	182	143	586	395	304	11		

Table A.4.2 Confusion matrix of vowels for CVSD.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1118	20	5				1							7	1		1152	97.0
ɑu	29	926	173							8	1	1		4		2	1142	81.1
ɑ	21	44	2226			1	2					8		1		1	2303	96.7
eː				##	1	4	6	2	62		1				3		1152	93.1
ɛi				8	1092	2	36							14			1152	94.8
øː				67	3	931	1		12	9	1	3	12	6	99		1144	81.4
ɛ			4	10	80	2	2114	6	43	1	3	10	3	17	9	2	2302	91.8
iː				16	3		2	954	80		32		54		10	1	1151	82.9
ɪ				58			32	46	991		4		5		15	1	1151	86.1
oː		7	13	4	1	6			1	981	8	134	2	1	2		1160	84.6
uː		1	1	5				63	36	17	934	18	50		21	2	1146	81.5
ɔ		1	23	1		1	10		5	87	15	##		2	3	2	1150	87.1
yː				7		11		208	41		101	1	708		71		1148	61.7
œy	5	5	6	2	33	14	16				4		4	1072	3		1164	92.1
ʉ			1	7		68	13	6	133	7	30	10	21		858	2	1154	74.4
conf	55	78	226	185	121	109	119	331	413	129	200	185	151	52	237	13		

Table A.4.3 Confusion matrix of final consonants for CVSD.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	816	139	165	7	1	16				1	5	2	1152	70.8
t	454	1521	260	24	9	14	3		1		14	4	2304	66.0
k	120	54	953	3	1	17				2		2	1152	82.7
f	134	116	124	451	167	135			4	9	9	3	1152	39.1
s	51	144	75	146	1787	70	1	1	6	4	14	5	2304	77.6
χ	28	17	78	45	38	909	2		2	1	29	3	1152	78.9
m	1	3				1	1475	101	672	44	6	1	2304	64.0
ŋ	1	1					160	777	197	11	3	2	1152	67.4
n	4	3			3		760	160	1330	29	7	8	2304	57.7
l	1	1	3	3	1	5	17	1	21	2065	81	105	2304	89.6
r	17	15	26	9	4	10	11	4	10	22	2172	4	2304	94.3
conf	811	493	731	237	224	268	954	267	913	123	168	139		

Table A.5.1 Confusion matrix of initial consonants for TU2400.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	773	117	114	22	3	50		30		22		1	2	1	8	2	5	2	1152	67.1
t	99	849	114	2	37	21	6	8	7	5			2			1	1		1152	73.7
k	58	45	1011	1	2	9	1	1		13				4		1	5	1	1152	87.8
b	62	5	8	653	81	8	1	6	1	3	11	9	13	9	261	11	9	1	1152	56.7
d	9	60	16	80	768	2		1	14			8	24	7	123	36	4		1152	66.7
f	76	12	26	6	1	658	12	288	1	57					4	1	10		1152	57.1
s	3	6	4			30	993	18	74	14						2			1144	86.8
v	35	14	13	30	24	282	13	383	37	46			14	10	196	24	31		1152	33.2
z	1	7	1	1	29	1	109	8	726	2	1	4	28	11	81	132	7	3	1152	63.0
χ	28	1	13			36	10	36	2	1011			1	6	5		10	1	1160	87.2
m	1			5		1	1			2	663	310	54	5	51	44	15		1152	57.6
n	2			2	5			2	1	3	85	824	85	9	39	85	9	1	1152	71.5
l	1				2	2		1	3	1	15	74	839	47	41	117	9		1152	72.8
r	5	5	12	10	12	2		4	3	15		12	69	847	98	10	46	2	1152	73.5
ʋ	11	2	4	78	7	3	1	4	1	1	27	11	37	32	873	32	27	1	1152	75.8
j		1	4	1	13		1		2		6	12	21	7	36	1029	18	1	1152	89.3
h	28	2	9	17	7	9	1	22	3	47	14	12	20	33	57	57	812	2	1152	70.5
conf	419	277	338	255	223	456	156	429	149	231	159	453	370	181	##	555	206	15		

Table A.5.2 Confusion matrix of vowels for TU2400.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1090	10	23			1		1				1		23	1	2	1150	94.8
ɑu	23	940	177							12				6	2		1160	81.0
ɑ	32	86	2148				5		1	2	1	19	1	6		3	2301	93.4
eː				1019	3	9	5	13	99		1				1	2	1150	88.6
ɛi				17	1069	1	34			2				29			1152	92.8
øː	1			23		930	8		12	2	2		17	2	143		1140	81.6
ɛ	3	3	5	54	61	6	1941	6	133		2	7	7	20	54	2	2302	84.3
iː				34	3	1	1	938	90		26	1	52		6		1152	81.4
ɪ			1	80	2	5	18	148	848		4	2	11		33		1152	73.6
oː		13	10	2	2	4			1	875	22	209		2	11	1	1151	76.0
uː		1		6		12		20	11	22	951	40	58		22	1	1143	83.2
ɔ		1	51	1			5	1	2	105	34	952	1		6	1	1159	82.1
yː				4		44	2	110	37		87	4	737	1	119	3	1145	64.4
œy	13	6	7		28	8	22	1	1					1055	2	1	1143	92.3
ʉ				3		103	12	3	67	1	12	10	50	4	903		1168	77.3
conf	72	120	274	224	99	194	112	303	454	146	191	293	197	93	400	16		

Table A.5.3 Confusion matrix of final consonants for TU2400.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	766	164	172	30	5	8	1				2	4	1152	66.5
t	131	1992	76	53	15	22	1	1		1	8	4	2304	86.5
k	71	87	936	23	4	22		1	1	1	6		1152	81.3
f	40	77	43	768	91	111	4	1	1	10	4	2	1152	66.7
s	4	44	4	73	2151	16				3	6	3	2304	93.4
χ	28	21	51	91	46	890	2		1	1	18	3	1152	77.3
m	5	4	4	2	2	1	1485	137	596	36	27	5	2304	64.5
ŋ	1	3	2				218	644	250	23	6	5	1152	55.9
n	7	4	1	3			712	169	1329	48	16	7	2296	57.9
l	8	5	6	5	1		36	7	33	2044	67	100	2312	88.4
r	11	10	10	22	6	18	8	4	14	24	2176	1	2304	94.4
conf	306	419	369	302	170	198	982	320	896	147	160	134		

Table A.6.1 Confusion matrix of initial consonants for US1200.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	767	120	127	31	2	32		17		8		1		3	9	2	33		1152	66.6
t	103	861	124	1	22	7	3	6	1	6		2		3		6	7		1152	74.7
k	52	57	973	3	5	8	1	2	1	34			1	2	2		11		1152	84.5
b	87	7	15	603	60	4		8	1	5	18	11	7	10	289	16	11		1152	52.3
d	17	64	21	71	732	5		1	5	4	3	8	9	18	118	60	16		1152	63.5
f	71	19	23	3	1	594	21	294	2	111			1		4		8		1152	51.6
s	4	4				48	931	42	87	29					1	1	1		1148	81.1
v	28	4	12	8	1	300	10	543	15	64	1	1	2	21	98	11	29		1148	47.3
z			2	1	12	4	173	15	807	2		4	2	15	24	84	7		1152	70.1
χ	13	3	12	2		28	9	35	6	1020		1		6	1		20		1156	88.2
m	4			9			1				589	416	32	9	62	17	13		1152	51.1
n	1										115	909	27	8	14	68	8	2	1152	78.9
l	4	1	1	6	6					2	16	161	703	52	54	113	31	2	1152	61.0
r	4	2	2	5	5	4		13	5	17	1	7	22	911	62	15	77		1152	79.1
ʋ	12		3	72	3	3		8	1	2	11	29	31	59	841	39	39	3	1156	72.8
j	1			4	7		1		6	1	4	15	17	8	32	1035	21		1152	89.8
h	27	4	15	16	2	12	1	10		50	15	18	5	32	58	38	849		1152	73.7
conf	428	285	357	232	126	455	220	451	130	335	184	674	156	246	828	470	332	7		

Table A.6.2 Confusion matrix of vowels for US1200.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1094	15	28				1							14			1152	95.0
ɑu	18	859	242		1		4			14		2		11		1	1151	74.6
ɑ	40	53	2136				9			4	4	42		8			2296	93.0
eː				986	3	4	11	21	115	2		1	1		2	2	1146	86.0
ɛi	7	1	1	61	1016		33	2	5					26			1152	88.2
øː		3		30	5	873	3		7	6	10	10	16	8	177		1148	76.0
ɛ	2		6	46	69	18	1876	7	188		3	2	5	24	73	1	2319	80.9
iː				21				997	62	1	29		40		1	1	1151	86.6
ɪ				77		9	26	168	798	2	8	5	13		38		1144	69.8
oː		10	13	2	2	4	1			874	32	211		1	6		1156	75.6
uː			1	2		3	2	29	3	19	965	57	55		11	1	1147	84.1
ɔ		1	31			1	2	3	1	87	43	970	1	1	11		1152	84.2
yː				6	1	30	4	88	14		160	2	806		41		1152	70.0
œy	23	11	5		45	28	25			3	4	1		1003			1148	87.4
ʉ			4	13	1	76	30	10	64	6	38	28	62		832		1164	71.5
conf	90	94	331	258	127	173	151	328	459	144	331	361	193	93	360	6		

Table A.6.3 Confusion matrix of final consonants for US1200.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	724	174	189	32		15	1		2	9	5	1	1152	62.8
t	173	1923	116	41	18	20	1	1	1	1	7	2	2304	83.5
k	83	66	930	8	1	34	1		4	7	18		1152	80.7
f	28	61	31	785	67	162			2	7	8	1	1152	68.1
s	1	10	5	67	2183	27			1	3	4	3	2304	94.7
χ	10	34	42	96	33	912	1			5	19		1152	79.2
m	8		4	3		4	1316	165	719	55	26	4	2304	57.1
ŋ	6	1	5			1	169	720	213	26	8	3	1152	62.5
n	4	9	1	5	2	5	725	218	1204	78	45	8	2304	52.3
l	3		5	2	1	5	39	9	29	2061	63	87	2304	89.5
r	10	5	31	24	2	27	14	1	11	22	2155	2	2304	93.5
conf	326	360	429	278	124	300	951	394	982	213	203	111		

Table A.7.1 Confusion matrix of initial consonants for LPC-10.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	753	124	94	27	19	43	1	21	1	13	1		3	2	19	2	28	1	1152	65.4
t	116	851	100	4	37	6	2	6	2	13	1				3	3	4	4	1152	73.9
k	132	85	835	6	6	21	2	17	3	18	2		1	3	2	3	15	1	1152	72.5
b	123	20	17	380	103	15	1	7	1	8	26	8	12	20	364	11	34	2	1152	33.0
d	37	88	45	74	664	6	1	2	4	8	2	20	12	16	128	30	15		1152	57.6
f	120	35	52	7	5	579	14	271	2	48		1	2	2	5	5	4		1152	50.3
s	3	6	1			41	975	38	67	8						3		2	1144	85.2
v	63	27	29	24	15	327	14	356	26	33	3	3	7	28	145	24	23	1	1148	31.0
z	2	4	2		8	16	250	12	702	13		3	9	32	33	62	3	1	1152	60.9
χ	77	28	44	3	1	169	45	84	3	658		1	2	7	1	9	28		1160	56.7
m	13	3	10	11	4	1		3		3	619	234	35	20	143	23	29	1	1152	53.7
n	5	4	6	5	18	3	2	2	3		174	716	70	10	52	69	12	1	1152	62.2
l	10	3	5	6	15	3	1	2	5	4	25	122	696	29	68	129	28	1	1152	60.4
r	26	11	26	30	30	16	3	6	4	33	12	8	46	715	96	19	69	2	1152	62.1
ʋ	57	7	12	69	30	12	1	9	3	4	33	27	22	63	734	26	45	2	1156	63.5
j	2	5	9	1	25	7	3	1	13	3	6	17	24	8	59	951	18		1152	82.6
h	61	12	45	12	12	47	2	30	6	84	15	18	16	69	85	43	595		1152	51.6
conf	847	462	497	279	328	733	342	511	143	293	300	462	261	309	##	461	355	19		

Table A.7.2 Confusion matrix of vowels for LPC-10.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	938	41	102	1	6		19			2	9		1	27	2		1148	81.7
ɑu	33	692	356			1	10			30	5	10		8		3	1145	60.4
ɑ	50	65	2012	3		4	17	7	2	22	29	87		4	6		2308	87.2
eː	1			990	14	3	25	21	89			4		1			1148	86.2
ɛi	13	1		87	927		86	5	14	1	3		2	8	4	1	1151	80.5
øː			1	45	3	815	11	2	19	21	4	5	40	24	174		1164	70.0
ɛ	2	1	13	61	98	6	1824	29	177	2	18	6	14	10	39	4	2300	79.3
iː				49	1	2	7	831	190		31	1	31		8	1	1151	72.2
ɪ			2	112	4	2	57	125	812	1	10	1	3	1	22		1152	70.5
oː		58	50	3	2	11	5	2		683	22	310		1	7	2	1154	59.2
uː			8	4	2	12	4	34	16	43	832	97	62	1	31	2	1146	72.6
ɔ	3	7	130	1	1	3	13	1	6	52	48	869	2		16		1152	75.4
yː			1	9	1	46	6	91	26		171	3	681	1	111	1	1147	59.4
œy	46	10	18	3	42	64	66	2	2	1	2	4	3	859	21	1	1143	75.2
ʉ			2	12		98	37	7	74	7	30	42	78		770	3	1157	66.6
conf	148	183	683	390	174	252	363	326	615	182	382	570	236	86	441	18		

Table A.7.3 Confusion matrix of final consonants for LPC-10.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	673	211	201	14	3	9	5	2	6	14	13	1	1152	58.4
t	229	1801	178	39	21	12	2	1	2	5	13	1	2304	78.2
k	109	105	877	12	4	18		4	2	5	16		1152	76.1
f	66	77	75	750	72	67	4	1	5	10	19	6	1152	65.1
s	2	9	3	147	2104	13			1	2	15	8	2304	91.3
χ	33	96	82	244	82	559		3	3	8	36	6	1152	48.5
m	16	21	19	9	4	11	1408	175	511	63	41	10	2288	61.5
ŋ	8	4	14	2		5	250	608	206	30	22	3	1152	52.8
n	10	16	9	10	4	13	826	212	1037	109	67	7	2320	44.7
l	13	11	9	3	2	14	67	8	36	1958	128	55	2304	85.0
r	40	47	69	47	11	34	20	12	18	38	1966	2	2304	85.3
conf	526	597	659	527	203	196	1174	418	790	284	370	99		

Table A.8.1 Confusion matrix of initial consonants for TU1200.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	761	116	103	14	8	51		42		18			3	1	12	5	18		1152	66.1
t	68	879	120	2	29	23	3	6		14		1		1	1		5		1152	76.3
k	62	67	929	2	4	18	1	21	2	25	4		2		1	3	14	1	1156	80.4
b	83	9	17	478	111	3		12	4	5	28	5	27	16	313	17	22	2	1152	41.5
d	14	69	35	71	708	1	1	1	13	7	4	18	22	18	99	69	1	1	1152	61.5
f	105	16	26	4	3	609	17	270	2	80			1		5	1	5		1144	53.2
s	3	9	3		1	47	928	18	101	17								1	1128	82.3
v	67	16	26	23	32	228	6	401	24	64	8	5	15	23	181	14	26	1	1160	34.6
z	1	5	8	4	68	8	122	17	703	8		11	26	23	55	101	14	2	1176	59.8
χ	19	9	27	4		62	7	43	2	942			4	15	2	4	8		1148	82.1
m	11	1	2	8	4	2		1			534	320	88	16	86	58	16	1	1148	46.5
n	4		1	3	4						91	769	114	13	36	107	13	1	1156	66.5
l		1		7	10	2			7	3	15	54	768	62	63	146	14		1152	66.7
r	13	8	4	19	7	1		7	5	24	7	14	94	766	94	21	66	2	1152	66.5
ʋ	28	3	5	77	7	9		5		7	41	29	76	59	730	46	30		1152	63.4
j	2	1	4	11	19			2	5	2	4	18	39	16	31	984	13	1	1152	85.4
h	23	11	24	8	5	16	2	21	4	61	30	24	37	37	84	66	698	1	1152	60.6
conf	503	341	405	257	312	471	159	466	169	335	232	499	548	300	##	658	265	14		

Table A.8.2 Confusion matrix of vowels for TU1200.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1089	14	33				2							4		2	1142	95.4
ɑu	21	823	268			1				15	2	2		7		1	1139	72.3
ɑ	50	97	2099		5	3	3			7	4	24	2	13	5		2312	90.8
eː	1		1	990	12	12	8	19	98				2		8	1	1151	86.0
ɛi	3	2	1	44	1004	3	57	1	6	1		2		28			1152	87.2
øː		1	1	19	1	936	13	2	19	5	1	3	16	13	133	1	1163	80.5
ɛ	7	3	7	79	118	20	1814	19	151		2	4	7	28	45		2304	78.7
iː			1	27	2	3	2	925	114		17	3	40		18		1152	80.3
ɪ				76	1	5	24	128	882		2	4	7		22	1	1151	76.6
oː		26	17	1	1	10	2		1	787	17	271	1	12	12	2	1158	68.0
uː			2	7		14	4	17	22	52	826	90	73	6	29	2	1142	72.3
ɔ	4	2	73			1	20	5	4	106	33	891		2	10	1	1151	77.4
yː			1	12		40	3	120	35		72	3	744	1	129		1160	64.1
œy	22	4	9		18	16	34			1	2		2	1039	1		1148	90.5
ʉ			2	6		121	24	14	79	1	13	13	65	2	808		1148	70.4
conf	108	149	416	271	158	249	196	325	529	188	165	419	215	116	412	11		

Table A.8.3 Confusion matrix of final consonants for TU1200.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	695	207	130	61	1	39	5	1	2	6	4	1	1152	60.3
t	142	1931	89	59	22	39	1		2	6	11	2	2304	83.8
k	82	101	893	29	1	24	3	1	1	3	13	1	1152	77.5
f	59	85	48	762	63	98	2	1	2	6	19	7	1152	66.1
s	3	69	6	93	2111	14			1		4	3	2304	91.6
χ	21	39	50	115	48	832	3	1	5	6	32		1152	72.2
m	26	13	5	8	1	2	1323	156	679	52	43	4	2312	57.2
ŋ		7	5	5	1	1	215	564	302	31	17	4	1152	49.0
n	11	7	3	12		3	720	232	1196	73	32	7	2296	52.1
l	9	9	12	7		4	54	12	36	2000	92	69	2304	86.8
r	8	12	21	33	7	29	20	8	27	64	2071	4	2304	89.9
conf	361	549	369	422	144	253	1023	412	1057	247	267	102		

Table A.9.1 Confusion matrix of initial consonants for FR1200.

	p	t	k	b	d	f	s	v	z	χ	m	n	l	r	ʋ	j	h	??	Tot	%
p	823	92	132	20	4	16		7		12	1	1		2	7	2	33		1152	71.4
t	114	861	118		18	5	1	3	2	6		2	2		1		18	1	1152	74.7
k	59	48	1009	2	2	3	1	2		6			1	2		3	14		1152	87.6
b	94	2	10	519	64	5		12	4	5	16	10	19	22	336	16	17	1	1152	45.1
d	18	76	24	44	692	4	3	7	18	11		13	13	21	128	59	20	1	1152	60.1
f	124	23	45	5		488	34	246	5	142	2	2	1	2	5	1	25	2	1152	42.4
s	13	5	11			15	968	9	112	7							2	2	1144	84.6
v	88	11	24	17	14	250	16	408	38	96	1	3	6	26	84	10	59	1	1152	35.4
z	1	3	5	3	8	2	179	18	795	7	1		7	13	20	81	11	2	1156	68.8
χ	21	5	39	1		22	11	28	3	1000				7	2	1	16		1156	86.5
m	6		1	3					2		508	374	75	13	87	64	18	1	1152	44.1
n	4	1	3	1	6				1	2	87	796	47	21	23	138	22		1152	69.1
l	1	1	1	6	2			1	9	2	8	97	712	73	57	145	36	1	1152	61.8
r	8	3	12	11	15		1	5	4	23	2	18	50	851	71	25	51	2	1152	73.9
ʋ	28		1	59	10	2		8	2	3	43	42	40	62	768	37	45	2	1152	66.7
j	3	2	9	5	19		2	4	7	1	2	38	16	9	21	974	40		1152	84.5
h	22	4	34	7	1	1		12	4	53	11	22	25	38	83	50	785		1152	68.1
conf	604	276	469	184	163	325	248	362	211	376	174	622	302	311	925	632	427	16		

Table A.9.2 Confusion matrix of vowels for FR1200.

	aː	ɑu	ɑ	eː	ɛi	øː	ɛ	iː	ɪ	oː	uː	ɔ	yː	œy	ʉ	??	Tot	%
aː	1080	14	32		3			1						20	1	1	1151	93.8
ɑu	41	772	290		8	1	2			20	7	5		9		1	1155	66.8
ɑ	25	66	2084		4	1	8	2	1	6	4	71	2	18	9	3	2301	90.6
eː				966	11	7	19	31	104	2	2	3	2		5		1152	83.9
ɛi	3		2	76	992		54		3	3				18		1	1151	86.2
øː		3	1	42	1	830	21	7	10	11	3	5	24	25	164	1	1147	72.4
ɛ		7	12	52	68	19	1782	10	204	1	3	3	16	25	88	6	2290	77.8
iː				51	1	2	8	922	106	1	11	5	29	1	14	1	1151	80.1
ɪ				108	2	6	69	135	786		11	1	11		31		1160	67.8
oː	1	44	26	4	2	6	2			752	36	277	1		4	1	1155	65.1
uː				7		5	7	39	24	30	823	110	70		33		1148	71.7
ɔ	5	3	55		2		16	7	8	70	49	924	4	1	8		1152	80.2
yː			2	9		41	5	123	26		116	6	727	1	94	2	1150	63.2
œy	20	9	19		23	45	34	1	2	5	1	5	4	975	5		1148	84.9
ʉ	1		5	5		97	26	20	92	1	24	23	75	10	776	1	1155	67.2
conf	96	146	444	354	125	230	271	376	580	150	267	514	238	128	456	18		

Table A.9.3 Confusion matrix of final consonants for FR1200.

	p	t	k	f	s	χ	m	ŋ	n	l	r	??	Tot	%
p	708	151	243	17		23	6			1	2	1	1152	61.5
t	133	1960	144	19	10	23		1		4	5	5	2304	85.1
k	80	60	963	11		14		1		1	21	1	1152	83.6
f	57	72	64	580	108	231	2		7	7	23	1	1152	50.3
s	2	31	6	39	2173	34				1	9	9	2304	94.3
χ	19	31	48	59	43	924	2		2	10	16	2	1156	79.9
m	10	3	3	5	1	6	1305	182	685	53	44	7	2304	56.6
ŋ	6	1	12	4	2	4	190	623	243	32	34	1	1152	54.1
n	5	7	6	3	1	8	729	194	1235	76	35	5	2304	53.6
l	9	7	10	5	1	5	42	17	35	2025	88	60	2304	87.9
r	10	5	34	11	1	27	14	5	18	50	2120	5	2300	92.2
conf	331	368	570	173	167	375	985	400	990	235	277	97		