ARTICULATION TESTS WITH A SIMULATED SUB-BAND CODER

M.A. SCOTT AND L.S. MOYE

STANDARD TELECOMMUNICATION LABORATORIES LIMITED

Articulation tests on simulated systems

Articulation tests at STL are normally done on telephone systems by our trained, female test team. The test material consists of random logatoms spoken in the carrier phrase KAN KON BYE ... ALSO: a complete test comprises 375 such phrases.

We required a comparison of the performance of several military transmission systems, including a computer simulation of a sub-band coder (SBC). This simulation runs in about 340 times real time, and the total computation time required to process the test speech material would have been about 350 hours. This represents around 30 days of overnight running on our mini-computer, so steps were taken to reduce the time needed.

Firstly, it is unnecessary to process every occurrence of the carrier phrase. Logatoms were read in groups of 15 with a single KAN KON BYE at the beginning of the group and a single ALSO at the end. Each logatom then acts as a carrier phrase for the next: this was successful since long pauses between logatoms for the benefit of listeners were not required. However, short pauses were required between logatoms (and between carrier phrase and logatoms) to eliminate coarticulation between them. The spoken material was recorded onto audio magnetic tape prior to being digitised into files on the computer's disk. The carrier phrase was edited out by hand into separate files for independent processing. Extraneous noises were also removed by hand at this point. Further savings were made by having a computer program scan through the speech data for substantial periods of silence and then remove these before the speech was put through the sub-band coder.

A simple test for silence was used as follows. Where necessary, the speech was passed through a 50 Hz comb filter to remove mains-related noise. A moving average of the signal was used to estimate and remove any DC offset. The RMS signal and RMS first difference were then calculated over short (10ms) frames. If both of these were less than fixed, pre-determined threshholds, the frame was a silence frame. If consecutive silence frames totalled more than 400 ms, they were considered to be a gap between logatoms. Except for 200 ms guard bands to prevent damage to the leading and trailing edges of the logatoms, such gaps were removed from the original recorded speech. As a check on this process, all material was listened to before it was submitted to the test team for evaluation: only one logatom was observed to have suffered even minor damage.

By these methods, the total computation time required for the simulations was reduced by about a factor five.

After processing through the SBC, a further program reassembled the processed speech and carrier phrase into the full, conventional test format (with a separate carrier phrase for each logatom), and the result was recorded onto audio magnetic tape for evaluation by the test team.

Unfortunately this saving could not be fully realised in terms of real (elapsed) time. Because a comparison of various systems was required, and because of the unconventionality of this test procedure, speech for every system (including real, non-simulated systems) had to be digitised and put through the same silence removal and reassembly process. This added considerable time in reconstruction (mostly due to disk seek times), and the net saving in total machine time was in fact a factor of about two.

Furthermore, the silence-detecting program, which also produced the data required for the reconstruction program, was not perfect. Its parameters were set to lose gaps in preference to damaging phonemes. In a few cases where the gap between two logatoms was particularly short, the logatoms were not separated at all. Thus there were a number of unseparated logatoms within a single carrier phrase. Such errors were corrected by hand (by splicing in a carrier phrase on the final audio tape) before passing the results to the test team.

One complete set of logatoms was passed through each of the six systems to be tested (Table 1). In view of the nature of the systems under test, male talkers were felt appropriate. Five were selected from the laboratory staff: each read 75 logatoms from each set to be processed through each system. Their speech was recorded via a telephone handset onto audio tape along with a high-quality (capacitor microphone) reference channel which was used as a check on the pronunciations of the inexperienced talkers. The speech was digitised off the audio tape through an anti-aliasing filter. For the V-S system, the speech was passed through the vocoder before filtering. All the data were compressed as described above. The two relevant tests were then put through the simulated sub-band coder and all were then reassembled and played back through the appropriate (real-time) system or systems and re-recorded onto audio tape. Because the computer was operating unattended, the playback level was set automatically by a programmable attenuator to give a constant rms signal over each group of logatoms. A single compromise level for all talkers was found by trial and error by playing back by hand a set of logatoms from each talker, and ensuring that the input level to the vocoder was within tolerance.

Articulation test results

The data from the listeners' and the talkers' sheets were processed by the normal computer program and the results are summarised in Table 1. This Table gives total percentage errors for initial, final and both consonants and vowel. The "total logatom" column gives the percentages of logatoms with at least one error.

Of the systems tested here, V, V-C and C-V had been tested previously with the conventional test format [1]. In each case, the error rate has risen greatly compared to the previous tests. A comparison of the control condition with other measurements [2] using conventional logatom tests on a bandlimited channel indicate a similar tendency. This is attributed to the lack of coarticulation in the present tests, which gives fewer clues to the identity of a phoneme than would otherwise be the case: it should be noted that the standard test purports to test single logatoms, but the logatoms are, of course, nearly always co-articulated with the carrier phrase. It appears that this is a much more severe test than one carried out in the usual manner and that results obtained using this modified test method must not be compared directly with those obtained by more normal means.

As expected, the vocoder on its own is (excluding the control condition) the best of the systems. Both S-V and V-S systems perform better than either C-V or V-C system: the performance improvement is, however, not as much as expected. It is noteworthy that the errors for "total logatom" are much more for the S-V system than for C-V, compared with the errors for the individual phonemes. Thus if a phoneme is heard incorrectly in a C-V system, it is much more likely that the remaining phonemes will also be heard wrongly than it would be for an S-V system. Errors in C-V systems tend to damage complete logatoms, whereas errors in S-V systems tend to occur in isolation.

Listening to the logatoms heard incorrectly by the test team showed that the disappointing performance of the S-V system was due largely to errors in the vocoder's voiced/unvoiced decision. This type of error is presumably particularly offensive to the ears of those, such as the STL test team, who are unaccustomed to the quality of vocoded speech. Indeed, in an informal test on several phonemes which none of the team had heard correctly, listeners more accustomed to the quality of ordinary vocoder speech had little trouble in hearing the logatoms correctly. This strongly suggests that evaluation of systems with peculiar quality should be made by the listeners who will use the system.

This work was done for the Joint Speech Research Unit under Government contract F7A/266/78.

References

1. Progress report for Government contract F7A/266/78, Jun 1980.
2. R. J. TAYLOR 1971 STL Internal Technical Memorandum 85 010. Articulation testing on reference working systems.

Table 1

percentage errors

system	initial cons.	final cons.	total cons.	vowel	total logatom
control	4	5	4	2	10
vocoder (V)	21	27	24	10	46
CVSD-vocoder (C-V)	40	46	43	19	69
SBC-vocoder (S-V)	32	38	35	14	62
vocoder-CVSD (V-C)	32	32	32	16	57
vocoder-SBC (V-S)	27	26	26	14	50
previous work: narrow band [2]	2.1	3.6	2.8	1.5	
vocoder [1]	17	16		9	33
CVSD-vocoder [1]	22	24		9	45
vocoder-CVSD [1]	21	26		12	47