Challenges in Using Audio Recordings for Large-scale Listening Comprehension Examinations

Keith Lenden-Hitchcock
Singapore Examinations and Assessment Board
Keith_John_Lenden-Hitchcock@seab.gov.sg

Abstract

Being a multi-cultural society, Singapore has a bilingual policy where students are required to learn English language as the first language and their own Mother Tongue Language (MTL) to as high a level as possible. A recent review by the Ministry of Education suggests the need to give greater focus on oral and listening skills in the learning of the MTL. Consequently, Oral and Listening skills are given greater importance in the MTL curriculum and examinations.

Each year, about 40,000 Grade 6 pupils sit a Listening Comprehension examination in the MTLs as well as in English Language. In these examinations, pupils listen to several audio recordings of passages before responding to a multiple-choice test. A key challenge for SEAB is to check each of about 2000 audio recordings. Checking each recording can take up to 40 minutes and therefore, the whole process can be very time consuming besides requiring a good number of staff involved.

Keywords: Listening Comprehension, Quality Checking, Compact Disk, Audio Recording

Focus of Paper

This paper shares some fresh approaches taken by the Singapore Examinations and Assessment Board to resolve a key problem on large scale listening comprehension examinations. In particular, it concerns the checking of the recordings in the shortest time possible with the minimum number of staff and more importantly, without compromising on the thoroughness of the checking process and the security of the examinations.

Background

In Singapore, every child has the opportunity to undergo at least ten years of general education. Almost all children start their formal education from age 6, spending six years in primary schools. At the end of Grade 6 they will sit their first major high stake national examination – the Primary School Leaving Examination, or PSLE - before proceeding to do a 4 or 5-year secondary education at the end of which they will sit the Singapore-Cambridge General Certificate of Education (GCE) Normal (Technical), Normal (Academic) or Ordinary Level examinations.

Bilingualism is a key feature of Singapore's education system. English is the main medium of instruction in school. At both the primary and secondary levels, pupils study various

subjects including the two required subjects English language and a Mother Tongue language. The bilingual policy is intended to equip the pupils with the language competencies to access both eastern and western cultures, and to develop a global outlook. These strengths give students a distinct competitive edge, helping them to appreciate their culture and heritage and connect with people from different backgrounds, so that they can thrive in a globalised world. Besides writing and reading skills, there is also strong emphasis in oral and aural skills in the learning of these languages.

Depending upon a pupil's ethnic group, his Mother Tongue may be Chinese, Malay or Tamil. While the vast majority of pupils would study the official Mother Tongue languages, Chinese, Malay or Tamil, a small minority is given the option to offer another available language in lieu of an official Mother Tongue language. The candidates may opt to be tested at one of three ability levels – *Foundation Level* for the less able, *Standard Level* for the majority and *Higher Level* for those studying for an in-depth understanding of the language.

The Listening Comprehension Examination

The PSLE English Language and Mother Tongue examinations comprise four papers testing writing skills, language use and comprehension skills, listening skills and oral communication skills.

Listening

In Listening Comprehension (LC) papers, candidates should be able to

- listen to a variety of spoken texts and demonstrate understanding of the content at the literal and inferential levels
- identify main ideas and recall details
- infer and draw conclusions
- follow a set of procedures[i]

Both the English and the Mother Tongue papers comprise 20 multiple-choice questions which test candidates' ability to understand the spoken language. The texts may be in the form of news items, announcements, advertisements, instructions, conversations, telephone conversations, speeches and short stories. In English language, graphic options are used for the first four items[ii].

A typical LC examination consists of a number of audio texts of between two and four minutes each and each text is followed by a set of two to five multiple-choice questions which are part of the audio recordings. These questions are also printed on the candidates' answer booklets and the candidates respond to these questions by shading the correct options on an Optical Answer Sheet (OAS). After a recording of a text and its accompanying questions are played to the pupils, the same text is played once more during which time the pupils may change or confirm their responses to the questions. This process is repeated for

each of the texts and questions. The examination last about 35 minutes for English language, 40 minutes for the Mother Tongue languages.

Ability Levels

For English language, a common set of texts and questions is used across all ability levels in the language for LC at the Grade 6 examination. For the three MTLs Chinese, Malay and Tamil, texts and questions are delivered at two levels – Foundation Level or Standard Level. Those pupils who study their Mother Tongue at the Higher Level are tested for LC Skills at the Standard Level.

Administration of the Listening Comprehension Examination

The listening examinations are administered on two distinct days – one for English language and the other for Mother Tongue language.

On the day of an examination, audio-recordings are dispatched to the schools, each school receiving two to twelve copies depending on the number of candidates in the school. In each school, candidates in groups of 30 or so wait for the teacher to bring the recording and administer the test. However, before the test can begin, the invigilator in charge of a particular examination room would have listened to the entire set of texts and questions usually an hour earlier, checking there are no flaws in the copy issued to him or her. Should an invigilator detect a flaw he or she would have to obtain another copy issued by the examining body.

Recording media

Until 1999, all audio recordings for the PSLE Listening Comprehension were made on audio compact cassette tapes. While audio CD technology has been available on the market since late 1982, it took a number of years before CD players were affordable (the first CD players retailed at about US$900[iii] – about US$1700 based on today's monetary value[iv]). Besides, compact-cassette tape recordings, especially voice-recordings where the hi-fidelity capabilities of audio CD are probably not required, were a proven and trusted technology since their introduction in 1963[v]. Worldwide, it was only in the early to mid 1990's that audio CD recordings became more popular than compact cassette recordings.

Tapes, of course, do have problems such as jamming or stretching requiring thorough checking procedures so as to minimise the probability of these problems occurring during an examination. However, such checking procedures could only ensure a tape would play well on a properly functioning cassette player. In addition it was not possible to check *all* cassette recordings, given the huge number required. Contingency plans which were in place to deal with problems which might occur on the day of the examination, were designed to solve those problems before the examination started and were very costly and labour-intensive.

In the year 2000, audio CD's, in place of cassette tapes, were used for the first time in the Listening Comprehension examinations in Singapore. While the possibility of faulty recordings was very much reduced with this technology, the need for checking could not be removed.

Checking the Quality of Recordings

In total, 12 Listening Comprehension Examinations are developed in various languages for the Grade 6 examination each year. After the master audio recordings of the examinations are produced, multiple copies of the recordings are replicated – more than 2000 in total.

Until recently, samples of 200 or so of these recordings were checked manually by up to 35 staff from SEAB, each spending approximately 5 or 6 hours per day over 4 weeks listening repeatedly to different copies of the same recording. Not only was the checking a very costly exercise in terms of manpower and security arrangements, it was an extremely frustrating and unpleasant task for all those involved. At times, an officer might have to replay a recording if he or she had lost concentration and was no longer listening actively!

Purposes of Checking

While only small samples would be adequate to check that a batch of recordings contain the correct content, a larger number of checked recordings is required should there be a need to replace a recording detected faulty by an invigilator before the examination started. When an invigilator finds a faulty recording he or she would have to travel to another nearby school used as a distribution centre to obtain a replacement which SEAB has to guarantee is in perfect working order. As such, all copies used for replacements were checked by SEAB.

Additional copies of checked recordings were also necessary for other uses such as when a pupil is unable to attend an examination centre (usually due to injury or illness) but is nevertheless well enough to take the examination. In this case an invigilator would travel to the individual's home - or sometimes a hospital – to administer the examination.

With cassette tapes, the need for such extensive checking was warranted but one might wonder if checking to the same extent is really necessary with Audio CDs. Nevertheless checking is still carried out, and for good reason.

Contingency for a Disaster

The severity and rapid spread of the SARS virus in 2003 resulted in many Singapore candidates having to be quarantined at home due to an infection of a close relative or exposure to an individual found to have contracted the virus. Existing contingency plans were not adequate should there be huge numbers of candidates unable to attend the Listening Comprehension Examination. While the near-pandemic situation had been fully contained by June of that year – well before the Grade 6 examinations in September – SEAB realised that

it would have to change its manual quality-checking procedures should there again be a suspected or actual pandemic. This would mean checking thousands, not just hundreds, of recordings – a technological approach is the only solution.

New Procedure of Checking

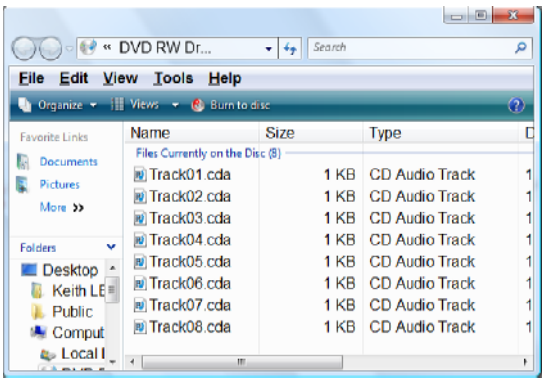
A new and easier checking procedure was therefore designed, taking advantage of digital recording technology. Unlike analogue copies of a recording (such as those found on audio cassette tapes) which are only approximate copies of an analogue master recording, digital copies of a digital master are *identical.* That being the case it should be relatively simple to use a computer to compare a CD-copy with the CD-master. A specially written program could thus be created to make the comparison without the need for the user to follow multiple steps or to spend hours listening to the full length of each recording repeatedly.

Such a program was developed after the 2003 near-pandemic SARS situation, though the programming involved turned out to be a little more complicated than anticipated. Nevertheless, the end result is that, not only can an individual now check 80 to 100 recordings in an hour (compared to 1 in 40 minutes), but CD reliability is guaranteed since the new system has zero probability of producing a false-positive verdict on an Audio CD. Further, more recordings can now be checked each year, requiring the services of only four or five officers for an hour or so each. An unsought-for benefit in the system is that only a few members of SEAB are exposed to the CD contents – only those directly involved in the examination. With the old method of checking, officers in charge of subjects such as Mathematics, Science and the Humanities were involved in addition to the Language officers.

In 2009, SEAB was well prepared in anticipation of an H1N1 outbreak, which fortunately did not manifest itself as the dreaded killer disease, when an additional 1000 recordings were both produced and individually checked in less than a day by just four officers.

Challenges in Using a Program for Checking

At first, one might assume that recordings on an audio CD are merely computer files just as one would find on a CD-ROM. After all, if one were to 'explore' an audio CD on a computer one might see something similar to the following:



If indeed, the 'files' Track*nn*.cda contained the actual tracks on the audio CD, checking would be straight forward since a program to compare a copied file with an original file is very elementary. However, the 'files' seen above do not really exist on the CD but are instead created by the operating system which obtains the information from an area on the CD known as the 'Table of Contents'. While these 'files' *can* be copied, they only contain information used by the CD drive to locate the beginning and ending of a track. They do not contain the actual sound data. Thus any program written to compare the tracks must be able to 'rip' the CD – that is locate the beginning of each track on the CD and read-off all the raw sound data until it reaches the end of that track.

Whilst commercially available programs can be used to rip a CD, it would be inconvenient for checking purposes since the user would have to initiate the ripping process, locate the ripped files on the computer and then initiate another program which performs a byte-by-byte check with the master tracks. In addition, the way raw sound data is stored on audio CDs makes this method unreliable in that there would be a large number of false-negative verdicts on the CDs which are checked. That is, two *audio* CD's (not PC data CDs) which have identical content would have a relatively high probability of being declared different. This appears especially true on machines such as laptop computers which may have lower-quality CD drives (perhaps due to the physical constraints on weight and size).

In order to understand why this is the case it is necessary to have knowledge of the way data is laid out on the surface of a CD.

CD Formats

The *physical* storage format for a CD is a little different from other disk devices. For example, while the information written to any disk device is stored in regions known as sectors, the physical layout of these sectors can differ. Mathematically, a sector is a portion of a circle between the centre, two radii and an arc, and indeed on many disk storage devices, data sectors are laid out in arcs inside the physical sector making it easy to calculate the position of any particular sector. This is not the case, however, for a CD where each sector is laid out end-to-end in a spiral about 5.7km long[vi] from the centre of the disk to its outer edge. On the CD itself, there are a number of different *logical*[1] formats in which data can be stored. Two of these formats are explained below.

> *CD Format for PC Data*
> For a PC data CD – that is, one for storing files - errors in data are not acceptable and therefore each 2352-byte sector is divided into 3 parts – a 16 byte header, 2048 bytes of user data and 288 bytes of error detection and correction code.

[1] I.E. the *way* data is stored in the physical sector.

CD Format for Audio Data

In the audio CD format, on the other hand, all 2352 bytes of a sector are used to store $1/75^{th}$ of a second of raw audio data - there is no header information nor is there any error detection and correction information within the sector.[2]

Reading a CD

To read a desired sector, the disk controller estimates where that sector begins, moves the read-head correspondingly, and starts reading from that point. For a CD meant for the storage of PC data, the disk controller can wait until the desired sector spins by since it can check the actual sector being read from information contained in the 16 byte header. For the audio CD, on the other hand, the disk controller simply reads from where it estimates a sector should be, even if it starts reading from the middle of a sector. The inability to read exactly at the desired point in an audio CD is known as *seek-jitter*.

Seek-Jitter Problems and False Negatives

Any program which does not take seek-jitter into account is likely to produce playback with audible clicks. For purposes of comparing one audio CD with another, it is essential that this phenomenon is correctly handled otherwise two identical CD's could be deemed different giving rise to a possibly high proportion of false negatives. Depending upon CD model and presumably its quality, the proportions of false negative comparisons due to seek-jitter were observed to be essentially as high as 100% on lower quality drives and virtually zero on higher quality ones.

Correcting for Seek-Jitter

Many commercially available ripping programs are able to make corrections for seek-jitter when a track is ripped from a CD. One of the ways such a program is able to do this is by deliberately reading overlapping regions from the disk, scanning the data to locate the *actual* overlaps, and removing all duplicated regions from the data. However, the exact beginning and ending of a track cannot be reliably determined since tracks are often padded at the beginning and end with silence[3] – this could still give rise to an alignment problem when comparing with another set of data. Further, reading overlapping regions causes some sectors to be read twice, somewhat slowing down the ripping process.

SEAB's disk checking software only re-reads data when seek-jitter is present. It is able to do this because, having a copy of the master recording stored in memory, it accepts only the expected data from the disk. When there is a mismatch, the program first checks to see if the

[2] In both these formats there is, however, a layer of error correction code at a sub-sector level. Some errors at this level are corrected or interpolated very efficiently by the CD hardware before being passed to the computer's processor.

[3] So that seek jitter is not an issue during normal playback as the person listening to a CD skips to his or her favourite tracks.

difference is due to seek-jitter and tries to correct for it. If it is unable to do so, the disk is rejected.

Disk Checking Procedure

Before disk checking can begin, the master CD from which all copies were manufactured is ripped and loaded into the disk checking program.

The program was designed so that minimal interaction between PC and user is necessary. Thus the user simply inserts a disk to be checked into the CD drive of the computer and closes the drive door - checking automatically starts as soon as the drive door is closed. After about 90 seconds the CD is automatically ejected from the computer, if, and only if, the CD it contains is identical to the master stored in memory. If the CD is different from the master, the CD is not ejected and a message is displayed on the computer screen indicating the cause of the difference.

In operation, a person can manage 5 machines simultaneously. After a little practice, it takes about two and a half to three minutes to fully process 5 CD's. This includes opening the original CD holders, inserting the disks into the drives, repackaging and sealing them. In an hour, the user can comfortably process 80 to 100 CD's if using desktop computers, but slightly longer using laptops.

Rejected Disks

In practice, SEAB has found that around 0.25% to 1% of CD's are rejected by the program. Inspection of these disks has sometimes shown that they may have been mishandled during packaging or when removing from the CD holder, though more often than not, there is no visible defect. On aurally checking the rejected CD's (the program informs the user the exact point of a mismatch), to date, no audible difference in the recordings has been found. However, there is a tiny difference in the bits extracted from the rejected disk, usually of duration no more than about 23 micro seconds and of magnitude only one part in 65536. Such errors are most likely due to microscopic defects on the surface[4] though we have found evidence of differences made by creating duplicates on different machines.

Apart from those disks where a visible defect is present, the other rejected disks, though truly differing from the master, are, for all practical intents and purposes, false negatives since the difference amounts to only one part in a 100 million of an entire 40 minute recording. This raises the issue of whether or not checking is absolutely necessary.

[4] On a PC data CD, such defects may be present too. However, the error detection and correction portion of a sector is able to rectify a number of such errors which cannot be rectified by the hardware.

The Necessity of Checking

Despite the small proportion of rejected disks, stringent quality control demands that SEAB delivers examination materials that are functional and 100% dependable. In Singapore, the PSLE is a very high- stakes examination as the posting of students to their *choice* secondary schools depends greatly on their results in the PSLE. It is not uncommon to hear that parents would book their leave one year in advance to take care and guide their children during the PSLE. As such, any slight irregularity in the examination would attract the media interest.

Program Development

The relative simplicity of the program's code and its use hinges on the exactness of copies of digital recordings. The first version of the program left no room for differences between master and copy, no matter how small those differences were. It was only during testing of the program on laptop computers, where an unreasonably high number of CD's were rejected, that SEAB realised CD-Audio data-retrieval is not as perfect as at first thought. Fortunately, by inspecting graphs of the sound waves, the programmer saw that two waves which should have been identical were in fact phase-shifted. Research on the internet as to its cause taught us the phenomenon of seek-jitter for which correction was built into the program.

The following year, tiny differences between CD and master were detected in some batches of CDs. Fortunately, the difference, besides being of negligible magnitude, was consistent across all CD's and thus correcting for this phenomenon was relatively simple.

The following year we detected three distinct batches of CD's, from one Mother Tongue, which should have been identical. While the difference was again negligible, it was very inconvenient for a person using the program to check the CD's since he or she would have to check against up to three different 'masters'. Thus an error margin was built into the checking so that inaudible differences between master and copy would be ignored; by default the program does not accept any differences, but the user can adjust a setting to change the default behaviour.

Most recently, a batch of CD's for one of the Mother Tongues gave inconsistent readings at approximately the same point in two of the tracks. Explicitly, the CD's were identical in every aspect excepting for about a quarter of a second. Further, each pair that was tested against each other was different to the other at those same points and the difference was of magnitude greater than the maximum setting the program allowed. The varying differences between each pair, and the fact that they occurred at the same points in the two tracks, suggest some flaw in the CD manufacturing or writing process. At the time of writing this paper, however, SEAB has yet to investigate the root cause of the phenomenon. As before the difference between master and copy was not audible – unless one knew exactly where to listen for the difference – but the anomaly proved very inconvenient and time consuming in an otherwise extremely smooth operation.

Conclusion

Extensive testing of the program on CD's over the years has shown that there are still instances of *effectively*[5] false negative comparisons. While most were due to technological and manufacturing factors that have been resolved, there are instances where the cause of the difference is, as yet, not understood by the system developers. To mitigate this, we provide more checked CDs to be used as standbys during the examination. Meanwhile, further testing is ongoing and there will be further refinement to the program so that inconsequential differences between master and copy do not result in unnecessary rejections.

References

[i] Singapore Examinations and Assessment Board. (2011). *PSLE English Language*, page 3. Retrieved from http://www.seab.gov.sg/psle/2011_PSLE_Subject_Info/0001_2011.pdf. See also
Chinese: http://www.seab.gov.sg/psle/2011_PSLE_Subject_Info/0005_2011.pdf
Malay: http://www.seab.gov.sg/psle/2011_PSLE_Subject_Info/0006_2011.pdf
Tamil: http://www.seab.gov.sg/psle/2011_PSLE_Subject_Info/0007_2011.pdf

[ii] Singapore Examinations and Assessment Board. (2011). *PSLE English Language*, page 4. Retrieved from http://www.seab.gov.sg/psle/2011_PSLE_Subject_Info/0001_2011.pdf

[iii] CED Magic. (nd). *Sony CDP-101 - The First Compact Disc Audio CD Player from 1982*. Retrieved from http://www.cedmagic.com/history/sony-cdp-101.html

[iv] 1 Soft Corporation. (nd). *Today's Dollars*. Retrieved on May 10th , 2011 from http://www.1soft.com/todaysdollars.htm

[v] Wikipidia, The Free Encyclopedia. (nd). *Compact Cassette*. Retrieved from http://en.wikipedia.org/wiki/Compact_Cassette

[vi] Terry McLean, Storage Device Articles. (2006). *CD ROM and CD R Tracks and Sectors*. Retrieved from http://e-articles.info/e/a/title/CD-ROM-and-CD-R-tracks-and-sectors/

[5] That is, while the rejected CD's have been found to be true negatives, the difference between these CDs and the master is usually inconsequential.