

Influence of compact disk recording protocols on reliability and comparability of speech audiometry outcomes: acoustic analysis

F DI BERARDINO, G TOGNOLA*, A PAGLIALONGA*, D ALPINI†, F GRANDORI*, A CESARANI

Abstract

Objective: To assess whether different compact disk recording protocols, used to prepare speech test material, affect the reliability and comparability of speech audiometry testing.

Material and methods: We conducted acoustic analysis of compact disks used in clinical practice, to determine whether speech material had been recorded using similar procedures. To assess the impact of different recording procedures on speech test outcomes, normal hearing subjects were tested using differently prepared compact disks, and their psychometric curves compared.

Results: Acoustic analysis revealed that speech material had been recorded using different protocols. The major difference was the gain between the levels at which the speech material and the calibration signal had been recorded. Although correct calibration of the audiometer was performed for each compact disk before testing, speech recognition thresholds and maximum intelligibility thresholds differed significantly between compact disks ($p < 0.05$), and were influenced by the gain between the recording level of the speech material and the calibration signal.

Conclusion: To ensure the reliability and comparability of speech test outcomes obtained using different compact disks, it is recommended to check for possible differences in the recording gains used to prepare the compact disks, and then to compensate for any differences before testing.

Key words: Audiometry; Speech; Hearing Level; Test Reliability

Introduction

Speech audiometry is an integral part of the audiometric test battery. Its diagnostic and therapeutic merits are recognised worldwide, mainly because hearing deficits are not always limited to an increased detection threshold for pure tones, but include other aspects of hearing such as distortion of sounds, comprehension of speech and noise discrimination. Much has been published on the outcomes of speech audiometry.¹ Some authors have suggested that it is preferable to pure tone audiometry in certain situations, such as the assessment of hearing difficulties in the elderly.²

Despite its great success, a number of issues deserve special attention in order to further improve the efficacy of speech audiometry. For example, much work has been done to establish clear criteria for the optimal choice of speech test material. However, little attention has been given to effective standardisation of protocols used during recording of speech material on compact disk (CD), and even less attention has been given to assessing

the possible effects that different CD recording protocols may have on speech test outcomes. This latter issue is typically underestimated as a marginal problem with little real impact on testing. Lack of clear standardisation of recording protocols, and the resultant variation in protocols, makes it difficult to compare test materials used in different laboratories, or even to compare different materials used within the same laboratory. Indeed, our current experience indicated that the use of widely accepted word lists is a 'necessary' but not 'sufficient' condition to enable full comparability of the outcomes of speech audiometric tests conducted at different laboratories. Furthermore, speech audiometry outcomes appeared to be influenced by the way in which these words lists have been recorded, and by the characteristics of the speech audiometers and other equipment used to play back speech material during testing.

This study aimed to assess and to quantify the extent to which different CD recording protocols, used to prepare speech material, affect the reliability

From the Audiology Unit, Fondazione Istituti di Ricovero e Cura a Carattere Scientifico (IRCCS) Cà Granda – Ospedale Maggiore Policlinico, and the Department of Otolaryngology, University of Milan, the *Istituto di Ingegneria Biomedica, Consiglio Nazionale delle Ricerche (CNR), Milan, and the †ENT Otoneurology Service, Scientific Institute S Maria Nascente, Don Carlo Gnocchi Foundation, Milan, Italy.

Accepted for publication: 20 January 2010. First published online 5 May 2010.

and comparability of speech test outcomes. To achieve this objective, we conducted acoustic analysis of a sample of CDs used for routine speech audiometry, to quantify any differences between CD recording procedures. This acoustic analysis focused on the measurement and comparison of the recorded levels of the calibration signal and the speech test material. Measurements were conducted of both the phono RCA (Radio Corporation of American) analogue output of the CD player, and in free-field, i.e., by measuring the sound pressure level of the calibration and speech test material as radiated from a loudspeaker located in the test room. Finally, in order to assess the possible impact of the different recording protocols on speech audiometry testing, the intelligibility of the tested speech materials was measured and compared in normal hearing subjects.

Materials and methods

Speech audiometry material

A sample of four different CDs was chosen from those used during routine speech audiometry in Italy. All these CDs contained phonetically balanced word lists currently in use for adult clinical testing. This material had been developed by Bocca and Pellegrini³ and successively modified by Turrini *et al.*⁴ and Todini.⁵ The material had been shown to give comparable results in normal hearing subjects; i.e. the word lists in the four CDs were of equal difficulty.^{4,5} Three of the four CDs contained speech material read by a man, while the fourth CD contained material read by a woman.

Acoustic analysis and measurement

For each CD, one single track consisting of one test list of spondees was acoustically analysed (20 spondees for CD1 and 10 spondees for CD2, CD3 and CD4). Each track contained intervals of a relatively constant duration (approximately 4 seconds) between successive test spondees. We also conducted acoustic analysis on each CD's calibration signal. Details of each CD's calibration signal are summarised in Table I.

Two different experimental arrangements were used for acoustic measurements.

In the first, the recording levels of the calibration signal and the speech material (i.e. the list of spondees) were measured directly from the RCA analogue output of a professional CD recorder (RW2000 CD recorder; Tascam, Montebello, USA;

total harmonic distortion during playback <0.004 per cent, frequency response 20–20000 Hz). The sound levels were measured using an audio analogue analyser Minilyzer ML1; NTI, Schaan, Liechtenstein; frequency response 10–20000 Hz, compliant with IEC (International Electrotechnical Commission) standard n. 61672, and were defined as the average root-mean-squared level expressed in dBu (re: 0.726 V_{RMS}).

The second experimental arrangement involved measurement of the integrated-averaging level over time of the sound pressure level (expressed in SPL (re: 20 µPa)) of the calibration signal and the speech material. Measurements were performed in free-field using an audio analogue analyser (NTI Minilyzer ML1), with a self-powered NTI MiniSPL microphone placed 1 m from the frontal speech loudspeaker (Genelek, Lisalmi, Finland; total harmonic distortion during playback <0.004 per cent, frequency response 20–20000 kHz). The audio analogue analyser used time weighting set to 'fast'⁶ and frequency weighting set to 'A'. The test room was compliant with international standard EN ISO 8252-2,⁷ and the environmental background noise was 40.5 dB SPL. The tested CDs' spondee lists and calibration signals were played through an audiometer (A177 Plus; Amplaid-Amplifon, Milan, Italy) connected to an external CD player (Tascam CD recorder RW2000) and to the loudspeaker (Genelek).

For each of the tested sounds, the minimum and the maximum sound pressure level values measured during the first 20 seconds of each signal were measured. For all CDs, the sound level measurements were done with the HL (Hearing Level) dial set to 60 dB HL.

Psychoacoustic evaluation

A group of 12 normal hearing volunteers (six women and six men; mean age 23 ± four years) were tested bilaterally ($n = 24$ ears) to determine their psychometric curves for the spondee lists of the four CDs. Tested subjects had no history of vertigo, balance disorders, hearing loss or otological problems. All subjects underwent otological evaluation (with otomicroscopy) and audiological evaluation (with pure tone audiometry and immittance audiometry), and obtained normal results. All subjects had a mean hearing level lower than 10 dB HL (at 0.5, 1 and 2 kHz) (mean value ± standard deviation, 7.2 ± 4.7 dB HL).

Psychometric curves were measured in a soundproof booth using TDH-49 headphones (TDH, Telephonics, New York, USA). The CDs were played through an audiometer (Amplaid A321) connected to an external CD player (Sony, Tokyo, Japan). Before performing the intelligibility test with each CD's selected spondee list, the audiometer was re-calibrated using the specific calibration signal recorded on that particular CD. Subjects listened to the four CDs in random order. Speech reception thresholds and maximum intelligibility thresholds were calculated from each subject's psychometric curve for each CD. Differences between speech reception thresholds and maximum intelligibility thresholds for the four CDs were tested

TABLE I

CALIBRATION SIGNALS RECORDED ON EACH CD

CD no	Calibration signal	Duration (s)
1	1000 Hz pure tone	60
2	992 Hz warble tone	33
3 & 4	Pink noise	30

CD = compact disk; no = number

using the Friedman test. When significant differences were found, post-hoc, multi-comparison analysis was performed using the Wilcoxon signed-rank test with Bonferroni's correction. Values of $p < 0.05$ were considered statistically significant.

Results and analysis

Table II shows the root-mean-squared levels of the calibration signal and the speech test material, as measured using the first experimental arrangement, for each of the four CDs. The last column shows the difference between the recording level of the speech test material and the calibration signal.

The calibration signal level differed among the four CDs, ranging from -13 dBu to 4 dBu. The level of the speech material (i.e. spondee list) was almost the same in all the CDs. The speech material level was not equal to calibration signal level in any of the CDs. In addition, the gain (i.e. the difference) between the speech material and calibration signal levels differed among the CDs, ranging from -11.0 dBu to 4.0 dBu. The measured gain between the speech material and calibration signal levels was positive in some CDs and negative in others. Specifically, in CD1 and CD2 the gain was negative (i.e. the calibration signal was recorded at a higher level than the speech material), whereas in CD3 and CD4 the gain was positive (i.e. the calibration signal was recorded at a lower level than the speech material).

Similar results were obtained with the second experimental arrangement, when measuring sound levels for the calibration signal and the speech material in free-field. The data shown in Table III parallel those shown in Table II, confirming that the speech signal was recorded at a different level to the calibration signal, in all CDs. In two of the four CDs (CD3 and CD4), the measured gain between the speech material and the calibration signal levels was as low as 1.5-3.0 dB SPL, whereas in the other two CDs the gain was greater (being -6.8 dB SPL in CD1 and -13 dB SPL in CD2; note that in these two CDs the gain was negative, i.e. the speech material was recorded at a lower level than the calibration signal).

Finally, the results of the psychoacoustic evaluation are illustrated in Figure 1, which shows subjects' mean percentage recognition scores as a function of speech level, using the spondee lists from the four

TABLE II

ROOT-MEAN-SQUARED LEVELS OF CALIBRATION SIGNAL AND SPEECH MATERIAL RECORDED ON EACH CD

CD no	Calibration signal (dBu)	Speech matl (dBu)	Δ_{dBu}^* (dBu)
1	-0.5	-7.0	-6.5
2	4.0	-7.0	-11.0
3	-13.0	-9.0	4.0
4	-12.0	-8.0	4.0

*Gain (i.e. difference) between the levels of the speech material (matl) and calibration signal. Gains and levels are expressed in dBu (re: 0.726 V_{RMS}). CD = compact disk; no = number

TABLE III

SPL LEVELS OF CALIBRATION SIGNAL AND SPEECH MATERIAL RECORDED ON EACH CD

CD no	Calibration signal (dB SPL)	Speech matl* (dB SPL)	$\Delta_{dB SPL}^\dagger$ (dB SPL)
CD 1	87.8	80.0-82.0 (81.0)	-6.8
CD 2	96.5	82.0-85.0 (83.5)	-13.0
CD 3	75.0	75.0-78.0 (76.5)	1.5
CD 4	75.0	73.0-83.0 (78.0)	3.0

*Data represent range of minimum to maximum peak levels (arithmetic mean peak levels). † Gain (i.e. difference) between mean speech level and calibration signal level. Gain and levels are expressed in dB sound pressure level (SPL) (re: 20 μPa). CD = compact disk; no = number; matl = material

CDs. The psychometric curves plotted in Figure 1 were found to be different for each CD. In particular, the curves obtained with CD1 and CD2 were below those obtained with CD3 and CD4. Although the four CDs contained similar speech material, and thus would be expected to give highly similar test outcomes, the actual experimental percentage recognition scores differed for each CD: at a given speech level, the percentage recognition scores obtained with CD1 and CD2 were always lower than those obtained with CD3 and CD4.

The observed differences in the psychometric curves for the four CDs were reflected in changes in both the speech reception threshold and the maximum intelligibility threshold. Figure 2 shows subjects' mean speech reception threshold and maximum intelligibility threshold values, for the four CDs. The mean speech reception threshold obtained with CD1 and CD2 was 16 and 18 dB HL, respectively, whereas that for CD3 and CD4 was 11 and 10 dB HL, respectively. Similarly, the maximum intelligibility threshold obtained with CD3 and CD4 was lower than that obtained with CD1 and CD2 (Figure 2).

For both the speech reception threshold and the maximum intelligibility threshold, the Friedman test indicated a significant difference between CDs ($\chi^2 = 23.91, p < 0.0001$). In particular, Wilcoxon signed-rank post-hoc analysis indicated that the speech

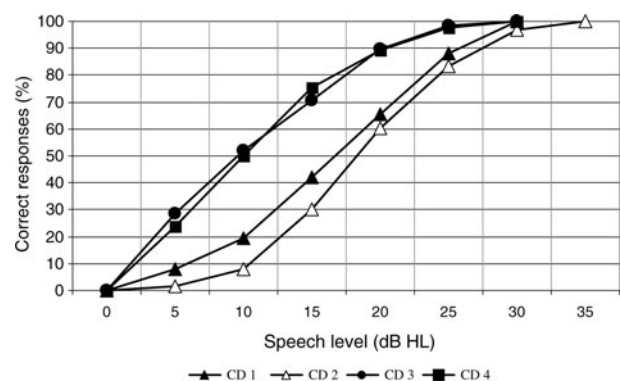


FIG. 1

Mean recognition percentage for 24 normal hearing subjects, for the different speech material recorded on each CD.

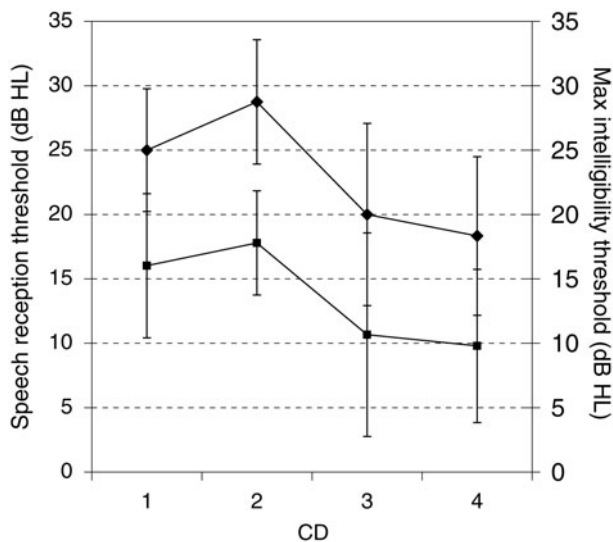


FIG. 2

Mean speech reception thresholds and maximum intelligibility thresholds for 24 normal hearing subjects, for the different speech material recorded on each CD. Whiskers indicate ± 1 standard deviation. Upper trace = maximum intelligibility threshold; lower trace = speech reception threshold

reception thresholds and the maximum intelligibility thresholds obtained with both CD1 and CD2 were significantly higher than those obtained with CD3 and CD4 ($p < 0.005$). It is notable that CD1 and CD2 contained speech material recorded at a considerably lower level than the calibration signal. No significant difference in speech reception threshold or maximum intelligibility threshold was found, comparing CD3 and CD4 (in which the speech material and calibration signal were recorded at reasonably equal levels). The maximum intelligibility threshold was within the normal range for all subjects, and for all CDs except CD2. When using the CD2 speech material, six of the 24 tested ears exhibited higher thresholds than normal.⁸

Discussion

The present study used acoustic analysis and psychoacoustic testing to evaluate the effect of different CD recording protocols on speech audiometry test outcomes. Acoustic analysis focused particularly on assessment of the level at which the calibration signal and the speech material had been recorded, for each CD. Of all the acoustic variables, it is well established that the recording level of the calibration signal, and its relationship to the recording level of the speech material, plays a critical role in ensuring the reliability and comparability of speech audiometry test outcomes.⁹ Because of the critical role of these two levels, a number of international standards have been published which establish specific requirements that the calibration signal and the speech material should fulfil. For example, the EN ISO 60645-2 standard¹⁰ (which deals with reference conditions for specification, testing and calibration of speech audiometers) is based on the assumption that the level of the calibration signal will be the

same as the average level of the speech material. A similar requirement is stipulated by EN ISO 8253-3,¹¹ another international standard addressing requirements for speech audiometry test material recording (see, also, the ANSI S3.6¹² standards and the recommendations of the American Speech-Language-Hearing Association¹³); specifically, standard EN ISO 8253-3¹¹ specifies that the level of the calibration signal should not deviate by more than ± 0.5 dB from the average level of the speech test material. Fulfilment of this requirement is necessary in order to guarantee that, when the speech audiometer output is set to 0 dB HL, the actual level at which speech test material is played back is the same, irrespective of the CD used.

Unfortunately, although the above-mentioned standards define, concisely and uniquely, the requirements for the average speech material level and the calibration signal level (i.e. they should be the same), there is no unique definition of the 'average level of speech'. For example, EN ISO 8253-3¹¹ allows the use of two different methods for measuring the level of the speech material: the 'equal speech level method' and the 'equal reference speech recognition threshold method'. The former method requires that the average speech level of each test word and test list be equalised with respect to the average level of all test speech material recorded on the CD; the latter requires that the average speech recognition threshold of each test word and test list be equalised with respect to the average recognition threshold of all speech test material recorded.

- **Speech audiometry is an integral part of the audiometric clinical test battery**
- **Little attention has been given to effective standardisation of the compact disk (CD) recording protocols used to prepare speech test material**
- **This paper assessed and quantified the effect of different CD recording protocols on the reliability and comparability of speech test outcomes**
- **Results from normal hearing subjects showed that the use of CDs containing similar speech material, but prepared with different recording protocols, leads to significantly different speech recognition thresholds and maximum intelligibility thresholds**

These two methods are only two examples of the techniques employed to equalise speech material. In practice, because of the lack of a clear and widely accepted definition of the average speech level, manufacturers use a variety of methods to equalise recorded speech material. The use of different methods for the equalisation of speech test material may lead to significant differences in the actual level of the recorded material, on different CDs.

This study's acoustic analyses revealed a significant difference in the actual level of recorded

speech material among different test CDs. In two CDs, the speech material was recorded at almost the same level as the calibration signal (the measured difference in levels ranged from 1.5 to 3.0 dB SPL). However, in the other two CDs assessed, the difference between the speech and calibration signal levels was very high: the speech material was recorded at a level 6.8 to 11.0 dB SPL lower than the calibration signal.

From a practical point of view, these findings mean that when the speech audiometer output is set to 0 dB HL, the actual level at which speech test material is broadcast differs from one CD to another. From a theoretical point of view, this difference in the relative gain between the reference and the speech test material could lead to different speech intelligibility thresholds: the higher the gain, the lower the expected threshold. The psychoacoustic results obtained from normal hearing subjects concur with this hypothesis. These results showed unequivocally that the psychometric curves, speech reception thresholds and maximum intelligibility thresholds obtained with speech test material recorded at a higher level than the calibration signal (i.e. with a positive gain) were significantly lower than these same parameters obtained with speech material recorded at a lower level than the calibration signal (e.g. results obtained with CD1 and CD2). Last but not least, three of the tested subjects exhibited higher than normal maximum intelligibility thresholds when tested with the speech material recorded on CD2, but were well within normal thresholds when tested with the other three CDs.

Conclusions

The results of this study confirm that different CD recording protocols have a real impact on speech test outcomes and on the comparability and reliability of speech audiometry. It is evident that correct calibration of the speech audiometer may not be sufficient to obtain comparable test outcomes if CDs prepared with different recording protocols are used. As emphasised by the EN ISO 60645-2¹⁰ and ANSI S3.6¹² standards and by the recommendations of the American Speech Language Hearing Association,¹³ in order to ensure the reliability and comparability of speech audiometry test outcomes obtained with different CDs, the sensitivity of the VU (Volume Unit) meter should be adjusted to compensate for the difference between the calibration signal and the speech material levels, before testing. The producer of the speech test CD should clearly specify how to modify calibration and test methods in order to equalise the calibration signal and the speech material. It is also recommended to experimentally check the relationship between the calibration signal and speech material levels, for example using an oscilloscope or a sound level meter.

Acknowledgements

The authors would like to acknowledge Mrs Isabella Bianchi at the University of Milan for her help in performing the speech audiometry tests on the normal hearing subjects analysed in this research.

References

- 1 Jerger J, Jerger S, Pirozzolo F. Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly. *Ear Hear* 1991;**12**:103–9
- 2 Martini A, Mazzoli M, Rosignoli M, Trevisi P, Maggi S, Enzi G *et al.* Hearing in the elderly: a population study. *Audiology* 2001;**40**:285–93
- 3 Bocca E, Pellegrini A. Statistical study of the phonetic composition of the Italian language and its practical application in audiometry with words [in Italian]. *Arch Ital Otol* 1950;**5**: 45–84
- 4 Turrini M, Cutugno F, Maturi P, Prosser S, Albano Leoni F, Arslan E. Bisyllabic words for speech audiometry: a new Italian material [in Italian]. *Acta Otorhinol Ital* 1993;**13**: 63–77
- 5 Todini L. Bisyllabic words for speech audiometry [in Italian]. In: Amigoni E, Todini L, Nume F, Del Bo L, eds. *Test di Valutazione della Percezione Uditiva*. Milano: Edizioni Aurion, 1997;2–48
- 6 European Committee for Electrotechnical Standardization (CENELEC). *Sound Level Meters (EN 60651:1994)*. Brussels: European Committee for Electrotechnical Standardization, 1994
- 7 European Committee for Standardization (CEN). *Acoustics – Audiometric test methods – Part 2: Sound Field Audiometry with Pure Tone and Narrow-Band Test Signals (EN ISO 8252-2:1992)*. Brussels: European Committee for Standardization, 1992
- 8 Becker W, Naumann HH, Pfaltz CR, Buckingham RA, eds. *Ear, Nose, and Throat Diseases: A Pocket Reference*, 2nd edn. New York: Georg Thieme Verlag, 1994
- 9 Brinkmann K, Richter U. Ensuring reliability and comparability of speech audiometry in Germany. In: Martin M, ed. *Speech Audiometry*, 2nd edn. London: Whurr, 1997;106–30
- 10 European Committee for Electrotechnical Standardization (CENELEC). *Audiometers. Part 2: Equipment for Speech Audiometry (EN 60645-2:1997)*. Brussels: European Committee for Electrotechnical Standardization, 1997
- 11 European Committee for Standardization (CEN). *Acoustics. Audiometric Test Methods – Part 3: Speech Audiometry (EN ISO 8253-3:1998)*. Brussels: European Committee for Standardization, 1998
- 12 American National Standards Institute. *Specifications for Audiometers (ANSI S3.6-2004)*. New York: American National Standards Institute, 2004
- 13 Calibration of Speech Signals Delivered Via Earphones. www.asha.org/policy [18 March 2010]

Address for correspondence:

Dr Gabriella Tognola,
Istituto di Ingegneria Biomedica CNR,
c/o Politecnico di Milano, Piazza Leonardo da Vinci 32,
I-20133 Milan, Italy.

Fax: +39 02 23993367

E-mail: gabriella.tognola@polimi.it

Dr G Tognola takes responsibility for the integrity of the content of the paper.

Competing interests: None declared
