**Acta Acustica united with Acustica**
**Vol. 98 (2012) 577 – 587**

DOI 10.3813/AAA.918539

# The Speech Intelligibility Aboard Metros in Different Running Conditions

Luigi Maffei[1)], Massimiliano Masullo[1)], Natalia Alexeeva[1)], Umberto Palmieri[1)], Vincenzo Paolo Senese[2)]

[1)] Built Environment Control Laboratory R.I.A.S., Second University of Naples, via San Lorenzo 81031, Aversa (CE), Italy. luigi.maffei@unina2.it

[2)] Department of Psychology, Laboratory of Cognitive Science and Immersive Virtual Reality, Second University of Naples, via Vivaldi 81100, Caserta, Italy

**Summary**

In this paper an investigation about the influence of noise on speech intelligibility (SI) inside metros under two different driving conditions (running in tunnel straight route (TS) and in curve (TC)), for female and male speakers with four voice levels (normal, raised, loud and shout) is presented. Perceptual laboratory tests were carried out where the noise of 11 metros was mixed with words of Italian language. The noise was always reproduced with the same level, while the speakers' voice amplitude differed, thus permitting to evaluate several signal-to-noise ratio (SNR) conditions to find out the SNR providing best intelligibility aboard. SI was quantified as the percentage of disyllabic words correctly understood (%WCU). Results showed, that %WCU was highly correlated with SNR. Poor intelligibility rates corresponded to "normal" voice condition. Fair SI can be guaranteed for SNR values greater than -6 dB in TC up to -3 dB in TS. Considering the use of raised vocal effort of 68,3 dB, the target noise level aboard, providing good SI, should be less than 71 dB. Besides, the results showed that the %WCU was influenced by the route conditions (driving in TC resulted in lower SI), the voice amplitude (%WCU rose with the increasing of the speech volume), the metro sound characteristics (SI scores differed in diverse metros), the gender of speakers (the voices of male speakers generally were more intelligible than female ones) and by the interaction of all the considered factors.

**PACS no. 43.50.Rq, 43.71.Gv**

## 1. Introduction

In the last decades the external noise of mass transportation systems has been well investigated in terms of exposure, annoyance [1, 2], psycho-physical and health effects [3, 4] in order to reduce its impact on inhabitants living in proximity to their routes. Likewise, satisfactory acoustical conditions have to be also guaranteed aboard these transportation means, representing environments where people spend a considerable part of their time travelling from and towards downtown. In a survey [5] on the national distributions of travel times to work in US population this time was estimated in about one hour per day, considering travel times to and from work.

The noise impact on passenger comfort inside standard and high speed trains was thoroughly studied [6, 7, 8, 9, 10, 11, 12, 13, 14], but as far as the acoustics aboard metros are concerned, very few investigations were accomplished [15]. Some recent measurements have showed that the equivalent sound level aboard often exceeds the value of 80 dB(A) [16, 17, 18, 19, 20].

Received 12 September 2011,
accepted 3 April 2012.

In EU States, in comparison with other mass transportation means, general preference towards the metros systems has been noticed in the recent years. Over the 1995-2004, the growth trend of metro use was observed [21], similar to the growth trend of private vehicles and even higher than buses/coaches and railway transport. In much the same way, a research of Italian Ministry of Infrastructures and Transport indicates that since 1990 till 2007 the number of passengers increased from 438 million to 738 million with an augmentation of 67.5% [22].

The combination of high noise levels, big number of passengers and considerable commuting time suggests to point out this topic.

The noise field inside metros can be characterized mainly by four types of sources:

- vehicle-related sources: traction engine, HVAC systems, sirens, others machinery aboard;
- interaction-related sources: rolling noise, squeal noise, breaking noise, tunnel crossing;
- transmission-related sources: airborne and structure-borne noise transmitted through the car boundary;
- human-related sources: activities inside wagons.

The noise emitted from the first three typologies of sources is non-stationary, it changes according to the operating conditions (engine RPM, accelerations, decelerations, curves, tunnel crossing, open/closed windows, etc.). Passenger activities during the trips are partially responsible for the loudness increase. These activities, consisting in speaking, using cell phones, headsets for music, portable gaming devices, strongly vary, depending on the social habits of populations. In many situations sellers, singers or musicians could constitute annoying noise sources.

During the years 2008 and 2009 several acoustic measurements and registrations have been performed in 17 European and not European metro lines in Naples, Rome, Milan, Turin, Paris, Berlin and New York. The results of this noise recordings campaign have demonstrated that often very high noise levels were noticed inside the metros during the rides [23]. These levels can cause stress effects, acoustical discomfort and, moreover, prevent the speech communication of passengers and require additional strenuous vocal efforts to reach fair intelligibility rates.

The Speech intelligibility (SI) is the principal measure for the assessment of human communication and communication systems efficiency [24]. It is the percentage of speech units that can be correctly identified by a listener over a given acoustic environment or the degree to which speech can be understood within given conditions [25]. Thomas Brand define SI as a "proportion of the speech items (syllables, words or sentences correctly repeated by a listener for a given speech intelligibility test)" [26].

SI can be determined using a variety of metrics. To provide measurement accuracy, reproducibility, validity and reliability, materials used for SI tests (sets of speech items such as nonsense syllables, single monosyllabic or disyllabic words or sentences) should to be homogeneous in SI and employ a representative sample of the speech sounds under the investigated conditions of speech communication, and the distribution of phonemes should represent the respective language. Just a possible cheaper cost of testing (including economy in using human subjects) and the potential for automation to simplify the administration of tests and analysis of the results should be considered [27].

A large number of SI computing methods is available, operating with both: technical (objective) and perceptual (subjective) procedures.

The methods of the first group use physical parameters to predict intelligibility according to a certain model: AI (Articulation Index) described in ANSI S3.5-1969 [28], SII (Speech Intelligibility Index) defined in ANSI S3.5-1997 [29], STI (Speech Transmission Index), RASTI (RApid Speech Transmission Index) defined in SS-EN 60268-16 [30], STIPA (Speech Transmission Index for Public Address Systems), %ALcons (Percentage Articulation Loss of Consonants), and speech clarity indexes.

The perceptual metrics for assessing SI, or subject-based methods, belong to the second category. These types of tests provide a more "realistic" measure of intelligibility because they use actual people and, in some cases, the participants are the final users of the systems being tested [24]. They use listeners with normal and sometimes impaired hearing to assess the effects of the talker and investigated environment on the clarity of the speech be means of PB (Phonetically Balanced Word Lists), MRT (Modified Rhyme Test), DRT (Diagnostic Rhyme Test) according to the ANSI/ASA S3.2-2009 [27], however many more tests have been carried out. Nonsense syllables and words are usually used for the analysis of information transmission, while sentences represent a communication situation more realistically.

Various factors affect the SI. First of all, to be intelligible in noisy conditions, the speech must have adequate audibility concerned with the signal-to-noise ratio (SNR), thus, speaker's signal level and background noise level are usually the main factors to be considered. SNR is the ratio of some measured aspect of a signal to a similar measure of concurrent noise expressed usually in a logarithmic form.

It is often seen that in noisy environment some speakers are definitely more intelligible than others. A significant variation in inter-speaker intelligibility is present even among speakers with similar SNR [31]. There were several attempts to analyze the speakers' gender impact on SI, however the results differ from research to research, and up to these days no unanimous consensus has been reached. In the earlier research Silversteina *et al.* [32] reported that male voices tended to be more intelligible than female voices. Similarly, Lees *et al.* [33] reported that gender of voice and the quality of signal do affect SI (regarding the Text-to-speech synthesis), and that generally the male voicing is more intelligible than the female voicing. On the contrary, Cerrato [34] and Barker and Cooke [35] have reported that women tend to be clearer than men. Instead, Tielen [36] indicated that no proper difference for male and female speakers was found, showing equal word and phoneme intelligibility under all noise conditions. At the same time, the differences between the individual speakers' intelligibility were rather large [35] where in the high noise conditions, listeners identified 24% of keywords from the least intelligible speaker but scored 68% for the most intelligible speaker.

There are also some interesting and contradictory results concerning the listeners gender impact on SI in the literature. Osafo-Yeboah *et al.* [37] and Ellis *et al.* [38] reported that in their researches they have not found any statistically significant difference in the intelligibility scores of male and female listeners, however they noticed that the overall impressions of speakers intelligibility differed. Women estimated the male speaker as more understandable while men indicated that the female speaker was more understandable [38]. On the contrary, Wilding and Cook (2000) reported that female listeners showed an enhanced ability to recognize female voices [39].

In this paper the research on the SI aboard metros is presented, the influence of different factors has been studied and comfort noise levels for fair SI are proposed.

## 2. Noise recordings

### 2.1. Investigated metros

Noise recordings was performed during the years 2008 and 2009 for 17 metro lines, having different constructive characteristics: tracks, wheels, speed, routes, etc. The investigated lines were: Milan (Line M1, Line M2, Line M3), Naples (Line M1, Line M2), Rome (Line MA, Line MB), Turin (VAL), Berlin (Line S3, Line U2), New York (Line M1, Line M4, Line MQ), Paris (Line M1, Line M6, Line M7, Line M12). The lines are supplied mainly with iron wheels except Turin VAL and Paris M1 and M6 which use rubber wheels. No information could be obtained in terms of effective metro speed, however the considered trains drive at a speed lower than 85 km/h, while the commercial speed (which includes the metro stops) ranges from 25 to 39 km/h.

### 2.2. Equipment setting and procedure of measurement

The binaural audio signals (16 bits/44.1 kHz) were recorded during the rides aboard metros with the use of a portable device "M-Audio Microtrack 24/96" and headphones “Sennheiser HDC 451”. Preliminary noise recordings were carried out in Naples, metro line 1, in order to check the dynamic range of the noise aboard. Sequences of recordings were arranged during travels on the entire route, to test different gain and level configurations for the digital portable device. The following gain and level settings were used as standard setup for on board metros recordings: Input levels=0 (L-R channels) and device gain=”low” (line-level). The left and right channel of the measurement chain were then calibrated recording a 94 dB/1 kHz pure tone signal of a 01dB-Metravib acoustic calibrator “CAL21”. The sound signals were then imported in the dBFA-Metravib software according to specific reference calibration factors.

For each metro line the records were carried out as far as possible in the same passengers’ position: in a motor wagon close to the doors. The recordings were performed on the entire routes for all the Italian metros and for more than 4 minutes for other metros. Surrounding conditions, such as tunnel crossing, acceleration/deceleration, breaking and open/closed windows were noticed during the travels or during the playback and analysis phases in laboratory.

### 2.3. Analysis of the physical data

The levels recorded by left and right channels of the headphones were averaged and analyzed in terms of $L_{eq,A}$, $L_{max}$, $L_{min}$, $L_{10}$, $L_{90}$ and in one third octave band. The overall (20 Hz–20 kHz) sound equivalent levels for each metro line and for the entire route (Figure 1) range from 65.1 dB(A) (Berlin S3) to 86,0 dB(A) (New York City M1). The level of 80 dB(A) is greatly exceeded inside all New York metro lines, Rome MB and Milan M3, characterizing high exposure conditions on board. Considering the $L_{max}$, very hard noise conditions for all New York lines and Naples M2, Rome MB and Milan M2 were noticed, although observing the $L_{10}$, these levels were present only for a short period. Nine of the seventeen lines showed values of $L_{90}$ greater then 65 dB(A). The quietest level was

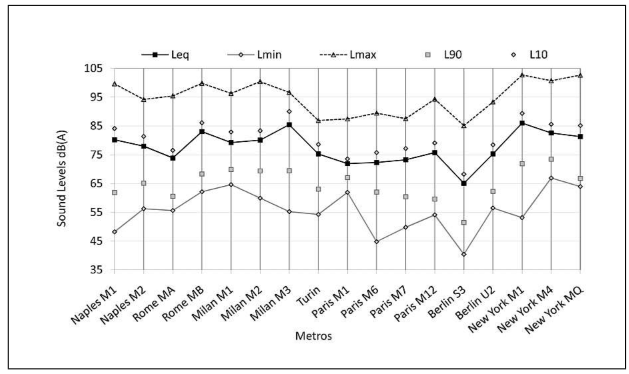


Figure 1. Sound equivalent levels of the metros.

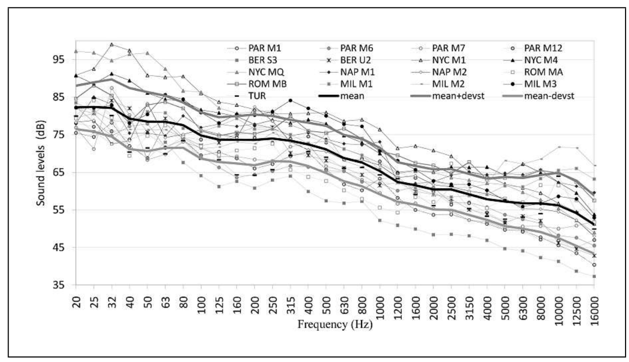


Figure 2. One third octave bands levels of all the metro lines, mean values and standard deviation range (grey area).

observed in Berlin S3, which route lies at open air and which has a long time of stops.

For every line, the signals were analyzed in one third octave bands (Figure 2). All the curves showed similar slope in the middle frequencies (about −4.0 dB for each frequency doubling) and concentration of energy in low frequency, especially for New York metros (M1, M4 and MQ).

Each route section between two successive stops was classified in terms of open-air or tunnel section and curve or straight section. Corresponding signals were extracted from total recordings and analyzed to obtain partial sound levels.

In Table I the values of the average sound levels $L_{eq,A}$ of each sub-section and doors closure (DC) are reported.

## 3. Perceptual tests

The perceptual tests for the evaluation of the speech intelligibility were performed in two running conditions: TS and TC.

### 3.1. Test materials and speakers

The American National Standard proposes three sets of test material in English for the SI measurement [27]: Phonetically Balanced Word List, Modified Rhyme test and Diagnostic rhyme test. There are no regulations available for the Italian language today, nevertheless, word lists

Table I. Overall sound equivalent levels in dB(A) on board. TC: "Tunnel Curve", OAC: "Open-Air Curve", TS: "Tunnel Straight", OAS: "Open-Air Straight", DC: Doors closure.

| City | ID | From | To | All | TC | TS | OAC | OAS | DC |
|---|---|---|---|---|---|---|---|---|---|
| Berlin | S3 | Hackescher | Zoologischer | 65.1 | - | - | 65.6 | 66.4 | 63.7 |
| | U2 | Ernst Reuter | Alexanderplatz | 75.3 | 79.6 | 75.5 | 70.5 | 70.5 | 74.9 |
| Milan | M1 | Bisceglie | Sesto Marelli | 79.2 | 78.6 | 80.1 | - | - | 83.2 |
| | M2 | Abbiategrasso | Cascinagobba | 80.1 | 83.8 | 75.2 | 70.9 | 72.4 | 76.1 |
| | M3 | Maciachini | San Donato | 85.4 | 91.4 | 84.5 | - | - | 77.5 |
| Naples | M1 | Dante | Piscinola | 80.2 | 86.1 | 78.8 | 83.4 | 76.5 | 84.0 |
| | M2 | Gianturco | Pozzuoli | 78.0 | - | 75.5 | 68.8 | 78.8 | 77.4 |
| New York | M1 | Chambers St. | 59th St. | 86.0 | 85.7 | 86.5 | - | - | 90.3 |
| | M4 | Brooklyn Hall | 14th | 82.6 | 83.6 | 80.2 | - | - | 88.7 |
| | MQ | 57th | Canal St. | 81.3 | - | 82.6 | - | - | 76.4 |
| Paris | M1 | Saint Paul | Châtelet | 71.9 | - | 71.6 | - | - | 77.9 |
| | M6 | Pasteur | Place d'Italie | 72.3 | 74.0 | 73.5 | - | 67.1 | 77.5 |
| | M7 | Place d'italie | Châtelet | 73.3 | 75.0 | 73.0 | - | - | 75.0 |
| | M12 | Concorde | Pasteur | 75.8 | 81.1 | 75.8 | - | - | 74.2 |
| Rome | MA | Anagnina | Battistini | 73.9 | 75.7 | 70.6 | - | - | 74.5 |
| | MB | Laurentina | Rebibbia | 83.0 | 87.8 | 81.6 | 78.6 | 71.8 | 85.0 |
| Turin | VAL | Porta Nuova | Fermi | 75.3 | 71.7 | 74.8 | - | - | 77.0 |

Table II. Example of word lists used for perceptual testing.

| A | B | C | D | E |
|---|---|---|---|---|
| cielo | nudo | rete | ira | fiele |
| era | quindi | campi | tarma | orlo |
| tordo | spinta | prova | chiesa | cento |
| alpi | giunco | tesa | unto | piedi |
| freno | sete | lunga | niente | tempo |
| chiuso | venti | bravi | zia | strada |
| sarti | lei | urli | gelo | mai |
| radio | seno | lire | scopa | calda |
| bionda | marzo | versi | ponte | onde |
| ali | sua | lega | neo | tela |

elaborated in 1950 by Bocca and Pellegrini are useful for intelligibility tests [40]. These lists, composed of 295 words and 50 logotomes, balanced by difficulty and phonetic composition, were used as speech materials for the perceptual tests. An example of list is presented in the Table II.

Four native and audiologically normal speakers (two women and two men) with a representative range of ages (25–45 years) were selected to describe different voice characteristics (young and mature voice, male and female speech). A list of 50 words composed by 5 sublists of 10 disyllabic words was assigned to each speaker and recorded. According to ANSI S3.2-2009 [27] requirements, before the recordings, the speakers were trained until they became thoroughly familiar with all test words and until they learnt to maintain a constant vocal effort during all the experiment. The records were performed in the anechoic chamber (5 m × 5 m × 5 m) of the Laboratory of the Second University of Naples. The receiver consisted of binaural headphones "HDS 451" on the head of a dummy,

Figure 3. Recording of SI test material: speaker-dummy distance 1 m face-to-face.

positioned face to face to the speakers, with a distance of 1 m from their lips (Figure 3).

Afterwards the single words extracted from the audio files were used to prepare the test signals consisting of 10 words, separated with a pause of 2 seconds from each other.

The comparison, in the octave frequency bands among 250 and 8000 Hz, between the recorded mean voice levels and those indicated in [29] is shown in the Table III.

### 3.2. Participants

The laboratory SI tests were performed with 56 Italian participants (n.32 male and n. 24 female) who stated that they had never had speech problems or faced hearing defects. Two groups of 28 listeners were chosen mainly from students and personnel of Faculty of Architecture of Second University of Naples. The mean age of the listeners was 28.5 years (s.d. = 8.1).

Table III. Sound levels of speakers versus ANSI Standard S3.5-1997 for different amplitude degree.

| Speaker | (+0dB) | (+6dB) | (+12dB) | (+18dB) |
|---|---|---|---|---|
| Giuseppe | 63.2 dB | 69.2 dB | 75.2 dB | 81.2 dB |
| Lia | 62.1 dB | 68.1 dB | 74.1 dB | 80.1 dB |
| Umberto | 64.5 dB | 70.5 dB | 76.5 dB | 82.5 dB |
| Valentina | 59.1 dB | 65.1 dB | 71.1 dB | 77.1 dB |
| Mean levels | 62.2 dB | 68.2 dB | 74.2 dB | 80.2 dB |
| | Normal | Raised | Loud | Shout |
| ANSI | 62.4 dB | 68.3 dB | 74.8 dB | 82.3 dB |

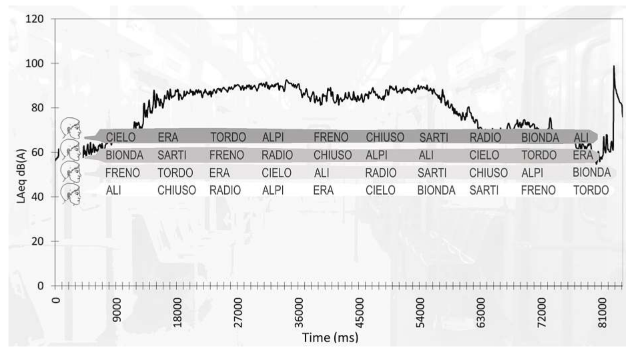


Figure 4. Example of the test time schedule.

Figure 5. Perceptual test in anechoic room.

## 3.3. Procedure

The tests were conducted in the anechoic chamber of the Laboratory of Second University of Naples. According to ANSI/ASA S3.2-2009 [27], before the experiment implementation all the listeners were trained until they could perform the appropriate perception-discrimination task, receiving in the beginning the following instructions (in Italian): "This test is aimed to define intelligibility of the speech aboard of metros in normal conditions of motion. Series of audio signals, consisting of background noise and sequence of words will be offered to listeners to understand. During this test 11 audio signals, recorded aboard of different metros, will be reproduced – one signal for each examined line. A list of 10 words is associated to every metro signal. Each listener is asked to identify and write down the words as he/she perceives them, considering that the same word list would be repeated with four different speech volumes, to make out which sound levels are more clear". The procedure among participants was randomized in terms of subway background sounds and of the word lists. Moreover, in order to avoid the participants' recalling of the words, for each subject and background sound condition the word lists were presented in a randomized order and always increasing the 4 speech volumes. An example of the test time schedule is shown on Figure 4.

A picture of a metro interior was put on the room wall to create an atmosphere close to that real, and soundtracks examples were given to listen to the people participating in the experiment as an additional training. The experiments were performed by personnel, completely familiar with system and qualified to work with all the laboratory equipment needed.

During the test the soundtracks, consisting of different metro TS and TC sections noise (only one type for each participant) and random sequence of words were reproduced by a laptop, a PCMCIA digital sound card and two loudspeakers "M 160 dB Technologies" (Figure 5). All the listeners accomplished their task individually (one by one) being positioned between the loudspeakers with a distance of 1.5 m from each sound source. Beforehand it was verified that the noise levels, reproduced at listeners' position, were similar enough (±1.0 dB) to those measured on board.

A person participating in experiments had to recognize speech cues byone speaker mixed with the sound of different metros.

The metro noise was held at the same reproducing level while the word lists were replayed with four amplitude degrees: +0 dB (normal), +6 dB (raised), +12 dB (loud), +18 dB (shout). The playback speech levels at the 4 vocal efforts are in line with the levels of the ANSI Standard S3.5-1997 [29], Pearsons [41] and with the more recently study of Cushing *et al.* [42]. Moreover, even if a loud or yelled (shouted) speech can lead to slightly differences in spectral composition [41] due to the different phonation process, an increase of the vocal effort leads to an increase in between-subject variation [42]. For this reason the authors have not considered any spectral changing as the vocal effort increase.

The listener wrote down the recognized words, completing a form, specially prepared to facilitate further elaboration and analysis of the results by statistical methods.

## 3.4. Results

The results of the word understanding for all the 56 participants (28x2) are summarized in figures 6 and 7, where the differences between all the metros for the four amplitude degrees are shown.

The results show poor intelligibility rates for "normal" voice condition: = 2.25/10 for TS and = 1.71/10 for TC conditions. Good words comprehension (= 8/10) is reachable with a +18 dB voice amplitude degree for 8 metros in TS and for 5 metros in TC.

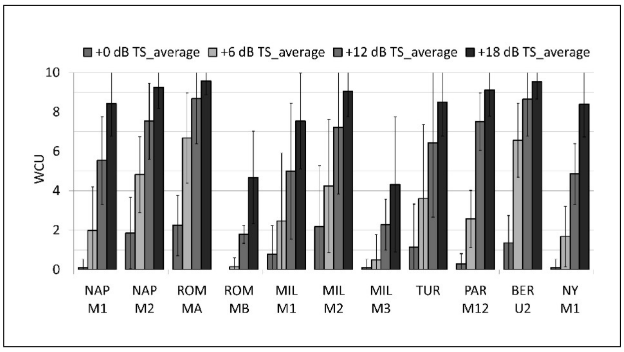


Figure 6. Number of words correctly understood for 28 subjects in TS condition. Four amplitude degree bars for all the metros and relative standard deviation (SD95%) are shown.

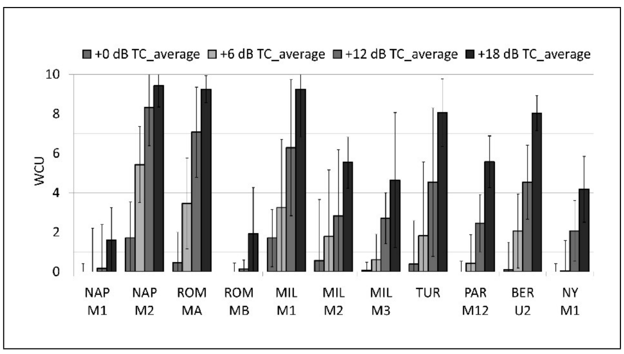


Figure 7. Number of words correctly understood for 28 subjects in TC condition Four amplitude degree bars for all the metros and relative standard deviation (SD95%) are shown.

Specific critical situations were detected in Milan M3, Rome LB for TS conditions, where less than 30% of words were recognized at +12dB. In TC with voice amplitude degree of +18 dB for 6 metros the number of words correctly understood were fewer than 6.

Afterwards the results of perceptual tests were compared with the real voice level of the speakers and of the metro noise.

## 4. Analyses of results

The analyses of the SI aboard metros were fulfilled comparing both physical and perceptual data in the octave band speech spectra between 250 and 8000 Hz. The characteristics of the 11 soundtracks, used for intelligibility tests, are reported in Table IV. The values, calculated by mean of dBSonic software were averaged for the left and right channels.

To investigate the effect of SPL and the specific psychoacoustic parameters on SI, scores were averaged over subjects as a function of each metro. Then, correlation analyses of acoustic parameters of each metro and mean intelligibility scores were carried out. The Pearson correlation coefficients were computed separately for each driving condition and for each voice level.

Results showed that in both TS and TC conditions and independently from the voice level there was a negative and significant correlation between Leq, Loudness and

Table IV. Characteristics of the metro noise soundtracks in the octave bands 250–8000 Hz. $L_{eq}$: Level in dB(A), $N$: Loudness (soneGF), $S$: Sharpness (acum), $FS$: Fluctuation Strength (cVacil), $R$: Roughness (cAsper).

| TS | $L_{eq}$ | $N$ | $S$ | $FS$ | $R$ |
|---|---|---|---|---|---|
| Naples M1 | 81.40 | 44.45 | 1.00 | 12.45 | 34.75 |
| Naples M2 | 74.05 | 36.25 | 1.10 | 4.90 | 35.45 |
| Rome MA | 76.70 | 39.05 | 0.95 | 19.35 | 27.20 |
| Rome MB | 87.70 | 72.75 | 1.00 | 10.10 | 40.60 |
| Milan M1 | 82.35 | 54.00 | 1.15 | 14.15 | 41.50 |
| Milan M2 | 75.95 | 38.45 | 1.15 | 5.05 | 34.95 |
| Milan M3 | 90.85 | 84.70 | 1.00 | 14.15 | 62.35 |
| Turin | 76.75 | 40.60 | 1.20 | 7.45 | 35.15 |
| Paris M12 | 74.20 | 36.20 | 1.20 | 5.75 | 37.55 |
| Berlin U2 | 74.30 | 34.15 | 1.20 | 7.70 | 36.30 |
| New York M1 | 85.35 | 55.05 | 1.00 | 11.80 | 45.25 |
| **TC** | $L_{eq}$ | $N$ | $S$ | $FS$ | $R$ |
| Naples M1 | 90.15 | 87.40 | 1.10 | 27.05 | 42.75 |
| Naples M2 | 74.60 | 36.20 | 1.00 | 9.40 | 29.75 |
| Rome MA | 77.00 | 43.85 | 1.30 | 20.25 | 34.45 |
| Rome MB | 90.80 | 89.85 | 1.30 | 19.05 | 51.25 |
| Milan M1 | 80.30 | 46.60 | 1.20 | 20.45 | 36.15 |
| Milan M2 | 83.10 | 64.50 | 1.65 | 20.75 | 42.70 |
| Milan M3 | 90.80 | 76.20 | 100 | 21.25 | 49.40 |
| Turin | 81.55 | 51.30 | 1.20 | 6.45 | 36.00 |
| Paris M12 | 85.45 | 62.05 | 1.00 | 27.85 | 40.30 |
| Berlin U2 | 84.25 | 63.60 | 1.05 | 19.30 | 38.15 |
| New York M1 | 88.80 | 95.40 | 1.00 | 27.00 | 48.50 |

Table V. Pearson correlations between metro noise characteristics and SI scores as a function of driving condition and voice level (N = 11). *: p-level < 0.05 (2-tailed); **: p-level < 0.01 (2-tailed).

| TS | Normal | Raised | Loud | Shout |
|---|---|---|---|---|
| $L_{eq}$ | -.462* | -.721** | -.872** | -.825** |
| Loudness | -.448* | -.698** | -.875** | -.896** |
| Sharpness | .110 | .270 | .413 | .375 |
| Fluct. Strength | -.069 | -.064 | -.220 | -.243 |
| Roughness | -.410 | -.605** | -.685** | -.710** |
| **TC** | Normal | Raised | Loud | Shout |
| $L_{eq}$ | -.561** | -.800** | -.794** | -.823** |
| Loudness | -.544** | -.761** | -.776** | -.838** |
| Sharpness | .064 | .055 | -.052 | -.014 |
| Fluct. Strength | -.358 | -.523* | -.483* | -.508* |
| Roughness | -.504* | -.734** | -.730** | -.777** |

Roughness and intelligibility scores. Only in TC condition in increased voice amplitude degrees there was a negative correlation between Fluctuation Strength and WCU. In general data showed that the correlations grow with the increase of the speech amplitude (Table V).

To analyze main and interactive effects of considered factors on intelligibility, a mixed factorial 4×11×2×2×2 ANOVA treated speech volumes and metros as four-level and 11-level within-subject factors respectively, and gender of speakers, gender of listeners and driving conditions as two-level between subject factors.

Results showed that SI is affected by speech volume, $F(3,144) = 1870.2$, $p < .001$, $\eta_p^2 = .975$, metro, $F(10,480) = 161.37$, $p < .001$, $\eta_p^2 = 0.771$, route condition, $F(1,48) = 29.04$, $p < .001$, $\eta_p^2 = .377$, gender of speakers, $F(1,48) = 15.97$, $p < .001$, $\eta_p^2 = .250$, and by the interaction of all the five considered factors, $F(30,1440) = 1.63$, $p = .017$, $\eta_p^2 = .03$. The Bonferroni correction has been used to analyze post-hoc effects.

As regards the speech volume, mean comparison showed that SI grows significantly as the speech volume increases ($Ms$ = 0.70, 2.47, 4.86, 7.10, respectively for the normal, raised, loud and shout voice volumes).

As regards the metro effect, mean comparison analyses showed that Metro 4 interfered particularly to the intelligibility, providing the lowest WCU scores ($M = 1.12$). Metros 7, 1 and 11 also interfered much to SI ($Ms$ = 1.94, 2.26, 2.64 respectively), Metros 9, 6, 8, 5, 10 provided medium SI ($Ms$ = 3.51, 4.17, 4.31, 4.53, 5.11 respectively) and, finally, Metros 3 and 2 showed highest SI scores ($Ms$ = 5.94, 6.07).

Mean comparison analyses of the speakers' gender SI revealed that men were more intelligible than women ($Ms$ = 4.36 and 3.20 respectively).

As regards the driving condition, mean comparison showed that SI was higher in Tunnel Straight ($M = 4.57$) than in Tunnel Curve ($M = 3.00$).

To interpret the interaction effect, further analyses have been carried out. The analyses were performed separately for the two driving conditions, and listeners' gender factor was excluded because it didn't show neither relevant main effect, $F(1,48) = .61$, $p = .437$, $\eta_p^2 = .013$, nor interaction.

### 4.1. Analysis of results for TS condition

As regards the TS condition, the mixed factorial 4×11×2 ANOVA that treated speech volumes and metros as four-level and 11-level within-subject factors respectively, and gender of speakers as two-level between-subject factor, showed that SI was affected by: speech volume, $F(3,78) = 1289.32$, $p < .001$, $\eta_p^2 = .98$; metro, $F(10,260) = 70.03$, $p < .001$, $\eta_p^2 = 0.73$; and by the interaction: speech volume×metro×speaker's gender, $F(30,780) = 4.34$, $p < .001$, $\eta_p^2 = 0.14$. The Bonferroni correction has been used to analyze post-hoc effects.

As regards the speech volume, mean comparison showed that SI increases significantly as the speech volume grows ($Ms$ = 0.9, 3.2, 5.9, 8.0, respectively for the normal, raised, loud and shout voice volumes).

As regards metro effect, mean comparison analyses showed that Metros 4 and 7 interfered more to the intelligibility, providing the lowest WCU scores ($Ms$ = 1.65 and 1.80 respectively), Metros 11, 5, 1 and 9 interfered less to SI ($M$s = 3.76, 3.95, 4.02, 4.87), then Metro 8 (Ms = 4.9), then Metros 6 and 2 ($Ms$ = 5.67, 5.87) and, finally, Metros 10 and 3 represented highest SI scores ($Ms$ = 6.53, 6.80).

As far as the three-way interaction effects in TS route condition are concerned, mean comparison between male and female speakers' intelligibility as a function of speech volume and metro (Table VI), showed that in some cases no statistically significant differences were noticed independently from the speech amplitude conditions (Metros 1, 2, 3,4, 9, 10 and 11). In other cases differences between male and female speakers SI were observed as a function of voice amplitude. When the speech volume was normal, the difference between male and female SI was observed in Metros 5, 6 and 8, where male speech tended to be clearer than the female one. For raised vocal effort the same direction of the SI difference was noticed in Metros 5, 6, 7, 8. In both loud and shout speech conditions, males were more intelligible than females in Metros 5, 7. Moreover, data showed that the sound pattern of the Metro 7, the noisiest one, interfered particularly with female voice intelligibility, which remained quite poor, regardless of the speech volume, while male speakers' intelligibility grew with the augmentation of speech volume.

### 4.2. Analysis of results for TC condition

As regards the TC condition, the mixed factorial 4×11×2 ANOVA that treated speech volumes and metros as four-level and 11-level within-subject factors respectively, and gender of speakers as two-level between-subject factor, showed that SI was affected by speech volume, $F(3,78) = 781,98$, $p < .001$, $\eta_p^2 = .97$, metro, $F(10,260) = 173.97$, $p < .001$, $\eta_p^2 = 0.87$, speakers' gender $F(1,26) = 17,14$, $p < .001$, $\eta_p^2 = .38$ and by speech volume×metro×gender of speaker interaction, $F(30,780) = 8.22$, $p < .001$, $\eta_p^2 = 0.24$. The Bonferroni correction has been used to analyze post-hoc effects. As regards the speech volume, mean comparison showed that SI increases significantly as the speech volume grows ($Ms$ = 0.5, 1.7, 3.7, 6.2, respectively for the normal, raised, loud and shout voice volumes).

As far as the metro effect in TC condition is concerned, mean comparison analyses showed that Metros 1 and 4 interfered more to the intelligibility, providing the lowest WCU scores ($Ms$ = 0.45, 0.52), Metros 11, 7, 9 and 6 interfered less to SI ($Ms$ = 1.59, 2.01, 2.12, 2.68), then Metros 8, 10, ($Ms$ = 3.705, 3.714), then Metros 3 and 5 ($Ms$ = 5.06, 5.13) and, finally, Metro 2 represented highest SI scores ($Ms$ = 6,223).

As regards the three-way interaction effects in TC route condition, mean comparison between male and female speakers' intelligibility as a function of speech volume and metro (Table VII), showed that in some cases no statistically significant difference, caused by speakers' gender, was noticed independently from the speech amplitude conditions (Metros 2, 3). In other cases differences between male and female speakers SI were observed as a function of voice amplitude. When the speech volume was normal, similarly to the TS condition results, the difference between male and female SI was observed in Metros 5, 6 and 8, where male speech tended to be clearer than female one. For raised vocal effort the same direction of the SI dif-

Table VI. The SI mean scores (WCU) comparison of the male and female speakers' for the TS condition. Note: equal letters indicate equal means ($p > .05$).

| | Voice amplitude | Normal | | Raised | | Loud | | Shout | |
|---|---|---|---|---|---|---|---|---|---|
| | Speakers | Male | Female | Male | Female | Male | Female | Male | Female |
| 1 | Naples M1 | 0.21$^a$ | 0.00$^a$ | 1.64$^a$ | 2.36$^a$ | 4.57$^a$ | 6.50$^a$ | 8.07$^a$ | 8.79$^a$ |
| 2 | Naples M2 | 1.43$^a$ | 2.29$^a$ | 4.43$^a$ | 5.21$^a$ | 7.14$^a$ | 7.93$^a$ | 9.07$^a$ | 9.43$^a$ |
| 3 | Rome MA | 2.29$^a$ | 2.21$^a$ | 6.57$^a$ | 6.79$^a$ | 8.50$^a$ | 8.86$^a$ | 9.36$^a$ | 9.79$^a$ |
| 4 | Rome MB | 0.00$^a$ | 0.00$^a$ | 0.14$^a$ | 0.14$^a$ | 1.43$^a$ | 2.14$^a$ | 4.43$^a$ | 4.93$^a$ |
| 5 | Milan M1 | 1.57$^a$ | 0.00$^b$ | 4.50$^a$ | 0.43$^b$ | 6.64$^a$ | 3.36$^b$ | 8.50$^a$ | 6.57$^b$ |
| 6 | Milan M2 | 4.21$^a$ | 0.14$^b$ | 6.21$^a$ | 2.29$^b$ | 7.71$^a$ | 6.71$^a$ | 9.07$^a$ | 9.00$^a$ |
| 7 | Milan M3 | 0.21$^a$ | 0.00$^a$ | 1.00$^a$ | 0.00$^b$ | 3.79$^a$ | 0.79$^b$ | 6.36$^a$ | 2.29$^b$ |
| 8 | Turin | 2.29$^a$ | 0.00$^b$ | 5.64$^a$ | 1.57$^b$ | 7.79$^a$ | 5.07$^b$ | 8.93$^a$ | 8.07$^a$ |
| 9 | Paris M12 | 0.43$^a$ | 0.14$^a$ | 2.79$^a$ | 2.36$^a$ | 7.43$^a$ | 7.57$^a$ | 9.00$^a$ | 9.21$^a$ |
| 10 | Berlin U2 | 1.71$^a$ | 1.00$^a$ | 7.21$^a$ | 5.93$^a$ | 8.64$^a$ | 8.64$^a$ | 9.57$^a$ | 9.50$^a$ |
| 11 | New York M1 | 0.64$^a$ | 1.21$^a$ | 2.14$^a$ | 1.21$^a$ | 5.29$^a$ | 4.43$^a$ | 8.14$^a$ | 8.64$^a$ |

ference was noticed in Metros 5, 6, 7, 8 and 10. In the loud speech volume condition the SI differences were detected in Metros 1, 4, 5, 6, 7, 8, 9, 10 and 11. In almost all listed cases the male speakers were more intelligible than the female ones. The only exception was Metro 1, where the female voice intelligibility was higher than the male one. Finally, as regards the shout condition, the same SI differences were detected in Metros 1, 5, 6, 7, 10 and 11. Also in this condition in Metro 1 the female voice SI was higher than the male one, and in all other metros men tended to be clearer than women. Similarly to TS conditions, the sound pattern of the Metro 7 interfered particularly with female voice intelligibility, which remained quite poor independently from the speech volume, while male speakers' intelligibility grew with the augmentation of speech volume.

The data show that speech intelligibility is influenced by the interaction between the considered components: noise type and sound characteristics, the speech reproduction volume and speakers' gender. It means that we should pay attention to the SI test correct implementation, using different types of voice and several noise patterns in the experiments, as to fulfill trustworthy measurements, all these factors should be manipulated. Only in this way it is possible to provide generalized and reliable data to define further appropriate recommendations or standards on internal comfort for the producers of transport means. In this sense, the measures of this study, since they used different sampling conditions, when all analyzed factors were manipulated, can be considered as quite reliable.

### 4.3. Analysis of results in terms of SNR

For the further analyses, the subjective speech intelligibility was quantified as the percentage of disyllabic words correctly understood (%WCU).

The Pearson correlation for the data of 11 metros in 2 driving conditions (TS and TC) for 4 voice levels and male and female speakers showed existence of the positive significant correlation between %WCU and SNR ($\rho = 0.827$, $p < 0.001$, $N = 352$).

The results of all perceptual tests for TS and TC conditions were fitted with a sigmoidal Gompertz function (Figure 8). The fitting curves show a slight difference between them for SNR less than about −10 dB where the difference in percentage of WCU is less than 5%. For values of SNR higher than −6 dB (TC) up to −3 dB (TS) an intelligibility percentage of 80% (Fair Intelligibility rating) can be guaranteed for the whole the conditions.

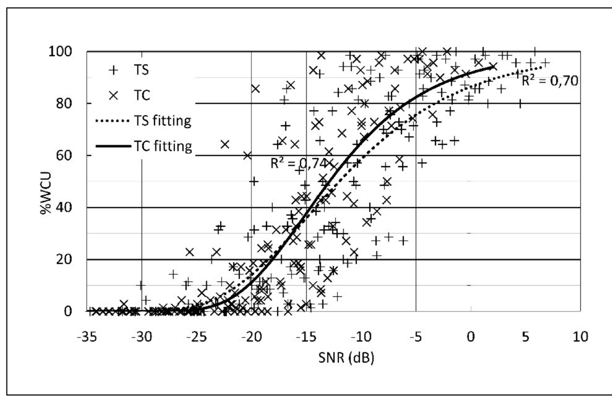


Figure 8. Perceptual intelligibility tests results for all the metro lines and speakers. "x": single answers for TC condition; "+": single answers for TS condition; solid curve: Gompertz fitting function for TC condition; dashed curve: Gompertz fitting function for TS condition.

Perceptual intelligibility tests results for all the metro lines, male and female speakers are represented on the Figure 9.

Considering the noise levels inside the metros (TS: from 74.9 to 95.2 dB, TC: from 75.5 to 96.9 dB) the 80% of WCU could be achieved only with an amplitude degree of speakers' voices increased of 12 dB or more (loud or shout condition) relatively to the basic level (Figure 10).

The same test results were subsequently analyzed in terms of speech intelligibility index SII. This index permits to get a good correlation with the speech intelligibility under a variety of adverse listening conditions. The input data: speakers' mean spectrum levels measured at 1 meter from their lips and the metro noises were elaborated according to the octave bands procedure as reported in ANSI

Table VII. Comparison of the male and female speakers' SI for the TC condition. Note: equal letters indicate equal means ($p > .05$).

| | Voice amplitude | Normal | | Raised | | Loud | | Shout | |
|---|---|---|---|---|---|---|---|---|---|
| | Speakers | Male | Female | Male | Female | Male | Female | Male | Female |
| 1 | Naples M1 | 0.00$^a$ | 0.00$^a$ | 0.00$^a$ | 0.00$^a$ | 0.00$^a$ | 0.36$^b$ | 1.14$^a$ | 2.07$^b$ |
| 2 | Naples M2 | 2.14$^a$ | 1.29$^a$ | 6.43$^a$ | 4.43$^a$ | 8.29$^a$ | 8.36$^a$ | 9.50$^a$ | 9.36$^a$ |
| 3 | Rome MA | 0.29$^a$ | 0.64$^a$ | 4.71$^a$ | 2.21$^a$ | 7.50$^a$ | 6.64$^a$ | 9.14$^a$ | 9.36$^a$ |
| 4 | Rome MB | 0.00$^a$ | 0.00$^a$ | 0.00$^a$ | 0.00$^a$ | 0.29$^a$ | 0.00$^b$ | 2.29$^a$ | 1.57$^a$ |
| 5 | Milan M1 | 3.43$^a$ | 0.00$^b$ | 4.79$^a$ | 1.71$^b$ | 8.29$^a$ | 4.29$^b$ | 9.86$^a$ | 8.64$^b$ |
| 6 | Milan M2 | 1.14$^a$ | 0.00$^b$ | 3.29$^a$ | 0.29$^b$ | 5.07$^a$ | 0.57$^b$ | 7.79$^a$ | 3.29$^b$ |
| 7 | Milan M3 | 0.14$^a$ | 0.00$^a$ | 1.14$^a$ | 0.07$^b$ | 5.14$^a$ | 0.29$^b$ | 7.14$^a$ | 2.14$^b$ |
| 8 | Turin | 0.79$^a$ | 0.00$^b$ | 3.57$^a$ | 0.07$^b$ | 6.64$^a$ | 2.43$^b$ | 8.79$^a$ | 7.36$^a$ |
| 9 | Paris M12 | 0.00$^a$ | 0.00$^a$ | 0.57$^a$ | 0.29$^a$ | 3.43$^a$ | 1.50$^b$ | 5.50$^a$ | 5.64$^a$ |
| 10 | Berlin U2 | 0.21$^a$ | 0.00$^a$ | 3.29$^a$ | 0.86$^b$ | 6.57$^a$ | 2.50$^b$ | 8.79$^a$ | 7.50$^b$ |
| 11 | New York M1 | 0.00$^a$ | 0.00$^a$ | 0.07$^a$ | 0.00$^a$ | 3.79$^a$ | 0.36$^b$ | 6.00$^a$ | 2.50$^b$ |

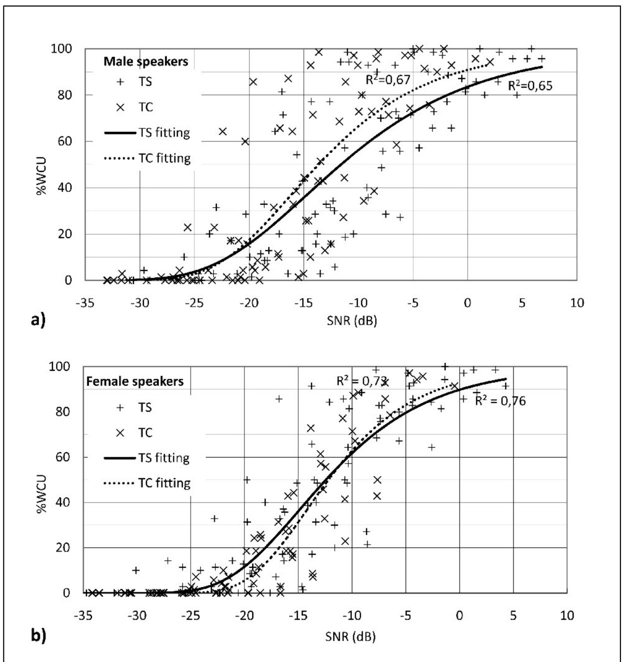


Figure 9. Speech intelligibility graphs for male speakers (a) and female speakers (b). "x": single answers for TC condition; "+": single answers for TS condition; dashed curve: Gompertz fitting function for TC condition; solid curve: Gompertz fitting function for TS condition.

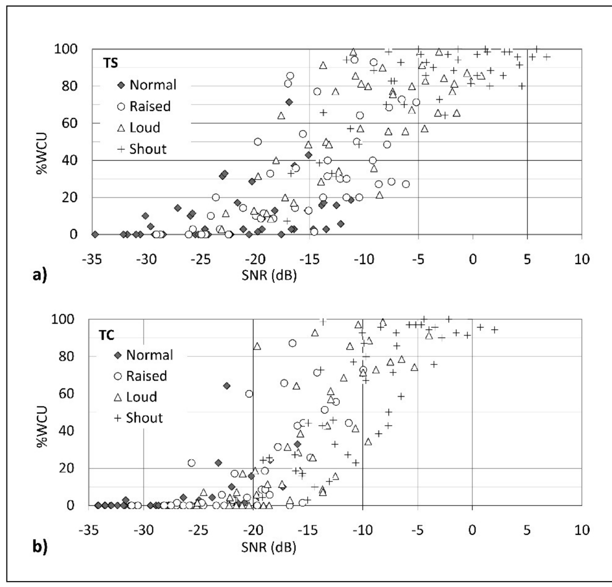


Figure 10. Percentage of WCU by listeners for all the metro lines and speakers per different amplitude degrees: TS condition (a) and TC condition (b).

S3.5-1997 [29]. The results are plotted related to the SNR (Figure 11).

Considering the SII benchmarks provided by [29] (poor < 0.45 and good > 0.75), the speech communication in the metro wagons is poor for almost all the conditions. Fair speech conditions are obtainable only with loud or shout vocal effort.

## 5. Conclusions

High noisiness aboard metros prevents normal conversation of passengers and requires additional strenuous vocal efforts to reach fair intelligibility rates. Perceptual laboratory tests were carried out to investigate the influence of noise on SI inside metros for two driving conditions: running in tunnel straight route and in curve, for female and male speakers with four voice levels (normal, raised, loud and shout).

Results showed, that intelligibility scores were highly correlated with signal-to-noise ratio. Poor intelligibility rates corresponded to "normal" voice condition. Fair SI can be guaranteed for SNR values greater than −6 dB in TC up to −3 dB in TS. Considering the use of raised vocal effort of 68,3 dB, the target noise level aboard, providing good SI, should be less than 71 dB.

Moreover, results showed that the %WCU was influenced by the route conditions (driving in TC resulted in lower SI), the voice amplitude (%WCU rose with the increasing of the speech volume), the metro sound characteristics (SI scores differed in diverse metros), the gender of speakers (the voices of male speakers generally were more intelligible than the female ones) and by the interaction of all the considered factors. It means that we should

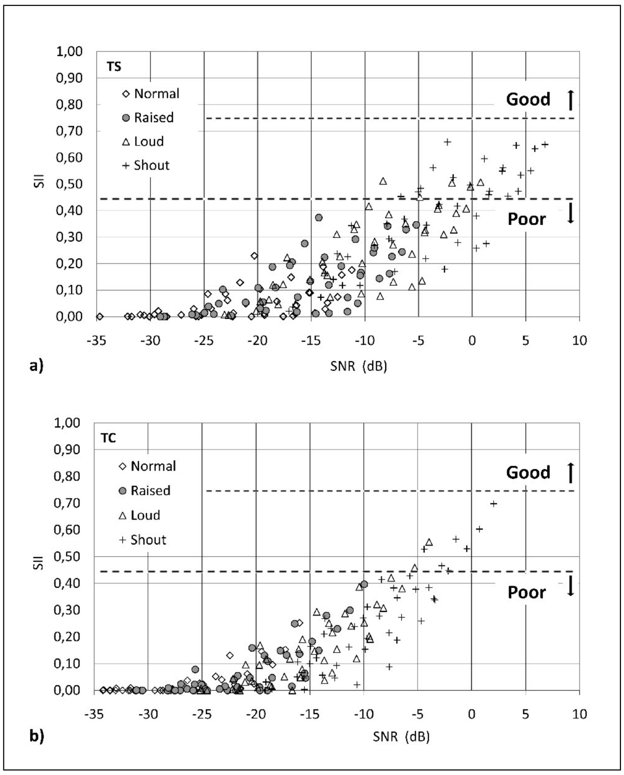


Figure 11. Speech Intelligibility Index SII for all the metro lines and speakers per different amplitude degree and operating conditions: TS condition (a) and TC condition (b).

pay attention to the SI test correct implementation, using different types of voice and several noise patterns in the experiments, as to fulfill trustworthy measurements, all these factors should be manipulated. Only in this way it is possible to provide generalized and reliable data to define further appropriate recommendations or standards on internal comfort for the producers of transport means. In this sense, the measures of this study, since they used different sampling conditions, when all analyzed factors were manipulated, can be considered as quite reliable. A possible future research, on the way to a prediction of preference for a male or female voice in a specific metro noise for the public announcement systems, the loudness patterns of the noises and the voices at different effort should be considered.

## References

[1] H. M. E. Miedema, H. Vos: Exposure-response relationships for transportation noise. Journal of the Acoustical Society of America **104** (1998) 3432–3445.

[2] H. M. E. Miedema, C. G. M. Oudshoorn: Annoyance from transportation noise: Relationships with exposure metrics dnl and denl and their confidence intervals. Environmental Health Perspectives **109** (2001) 409–416.

[3] H. M. E. Miedema, H. Vos: Associations between self-reported sleep disturbance and environmental noise based on reanalyses of pooled data from 24 studies. Behavioral Sleep Medicine **5** (2007) 1–20.

[4] W. Babisch: Updated review of the relationship between transportation noise and cardiovascular risk. Proceedings of the 12th International Congress on Sound and Vibration, Lisbon, 2005.

[5] B. McKenzie, M. Rapino: Commuting in the united states: 2009. American Community Survey Reports, U.S. Department of Commerce Economics and Statistics Administration U.S. CENSUS BUREAU, issued September 2011.

[6] A. E. Hardy: Measurement and assessment of noise within passenger trains. Journal of Sound Vibration **231** (2000) 819–829.

[7] Y. Soeta, R. Shimokura: Noise and sound field measurement in train vehicles. Internoise 2008, 37th International Congress on Noise Control Engineering, Shanghai, China, 26-29 October 2008.

[8] M. S. Khan: Evaluation of acoustical comfort in passenger trains. Acta Acustica united with Acustica **88** (2002) 270–277.

[9] M. S. Khan: Effects of masking sound on train passenger aboard activities and on other interior annoying noises. Acta Acustica united with Acustica **89** (2003) 711–717.

[10] Y. Soeta, R. Shimokura: Evaluation of speech intelligibility of sound fields in passenger train compartments. Acoustical Science and Technology **30** (2009) 379–382.

[11] F. Led Tourneaux, S. Guerrand: Assessment of the acoustical comfort in high-speed trains at the SNCF: integration of subjective parameters. Journal of Sound and Vibration **231** (2000) 839–846.

[12] S. Kuwano, S. Namba, T. Okamoto: Psychological evaluation of sound environment in a compartment of a high-speed train. Journal of Sound and Vibration **277** (2004) 491–500.

[13] C. Patsouras, H. Fastl, U. Widmann, G. Hoelzl: Psychoacoustic evaluation of tonal components in view of sound quality design for high-speed train interior noise. Acoustical Science and Technology **23** (2002) 113–116.

[14] W. Wei, A. Bockstael, B. De Coensel, D. Botteldooren: Effects of Chinese high-speed train compartment noise and speech on task performance. Proceedings of Forum Acusticum, 27 June – 1 July 2011, P:4.

[15] N. Yoshihisa, M. Fukai, T. Yamaguchi: Noise levels inside subway cars in Western Europe, U.S.A. and Japan. Journal of Acoustical Society of Japan **24** (1968) 69–75.

[16] M. E. Aly: Noise assessment inside the greater Cairo underground second-line metro. Sādhanā **30** (2005) 47–55.

[17] R. M. Gershon, R. Neitzel, M. A. Barrera, M. Akram: Pilot survey of subway and bus stop noise levels. Journal of Urban Health: Bulletin of the New York Academy of Medicine (2006).

[18] L. De Gregorio, L. Maffei, M. Masullo: Esposizione dei passeggeri a rumore: Il caso della linea metropolitana di Napoli. 34° Convegno Nazionale dell'Associazione Italiana di Acustica AIA, Firenze, 13-15 giugno 2007.

[19] L. Maffei, M. Masullo, U. Palmieri: Comfort acustico a bordo di metropolitane. 36° Convegno Nazionale dell' Associazione Italiana di Acustica, Torino, 10–12 Giugno 2009.

[20] H. Erol: Measurements and analysis of the noise to which passengers are exposed in Istanbul metro. 16th International Congress on Sound and Vibration ICSV16, Kraków, Poland, 5-9 July 2009.

[21] MVV Consulting: Tractebel development engineering. Preparation of a green paper on urban transport: report on urban transport in Europe, 2007.

[22] Ministero delle Infrastrutture e dei Trasporti, Dipartimento per i Trasporti Aereo, Marittimo, per gli Affari Generali, Conto Nazionale delle Infrastrutture e dei Trasporti, anni 2007–2008, Sistema Statistico Nazionale, 2008.

[23] M. Masullo, L. Maffei, N. Alekseeva, U. Palmieri: An investigation of the speech intelligibility on board metros. 8th European Conference on Noise Control EURONOISE, Edinburgh, Scotland, October 26-28 2009.

[24] M. A. Blue, C. Ntuen, T. Letowski: Speech intelligibility measured with shortened versions of callsign acquisition test (CAT). Applied Ergonomics **41** (2010) 291–294.

[25] T. Letowski, R. Karsh, N. Vause, R. Shilling, J. Ballas, D. Brungart, R. McKinley: Human factors military lexicon: Auditory displays. U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Grounds, MD, technical report, 2001.

[26] D. Havelock, S. Kuwano, M. Vorländer: Handbook of signal processing in acoustics. Part II. 2009. 197-204.

[27] ANSI: American national standard method for measuring the intelligibility of speech over communication systems. American National Standard Institute, ANSI/ASA, New York, S3.2-2009.

[28] ANSI: American national standard methods for the calculation of the articulation index, ANSI S3.5-1969. American National Standard Institute, New York, 1969.

[29] ANSI: American national standard methods for the calculation of the speech intelligibility index, ANSI S3.5-1997 (Revision of ANSI S3.5-1969, Reaffirmed in by ANSI in May 18, 2007). American National Standard Institute, New York, 2007.

[30] SS-EN 60268-16: Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index. SIS – Swedish Standards Institute, Stockholm, Sweden, 2003.

[31] V. Hazan, D. Markham: Acoustic-phonetic correlates of talker intelligibility for adults and children. J. Acoust. Soc. Am. **116** (2004) 3108–3118.

[32] B. Silversteina, R. C. Bilger, T. D. Hanley, M. D. Steer: The relative intelligibility of male and female talkers. Journal of Educational Psychology **44** (1953) 418–428.

[33] N. Lees, C. Stevens, J. Vonwiller: Implicit measurement of intelligibility of male and female voice text-to-speech (TTS) synthesis in noise using a phoneme detection task. Proceedings of the 9th Australian International Conference on Speech Science and Technology, Melbourne, Australia, December 2 to 5, 2002.

[34] L. Cerrato: Using a multitalker database to investigate the speaker-related characteristics of speech intelligibility. NAL. Annual Report, 1995.

[35] J. Barker, M. Cooke: Modelling speaker intelligibility in noise. Speech Communication **49** (2007) 402–441.

[36] M. T. J. Tielen: Male and female speech. An experimental study of sex-related voice and pronunciation characteristics. PhD thesis. In the proceedings of the Institute of Phonetic Sciences Vol. 16, University of Amsterdam, Amsterdam, The Netherlands, 1992.

[37] B. Osafo-Yeboah, X. Jiang, M. McBride, D. Mountjoy, E. Parka: Using the callsign acquisition test (CAT) to investigate the impact of background noise, gender, and bone vibrator location on the intelligibility of bone-conducted speech. International Journal of Industrial Ergonomics **39** (2009) 246–254.

[38] L. Ellis, L. Reynolds, D. Fucci, B. Benjamin: Effects of gender on listeners' judgments of speech intelligibility. Perception and Motor Skills **83** (1996) 771–775.

[39] J. Wilding, S. Cook: Sex differences and individual consistency in voice identification. Perceptual and Motor Skills **91** (2000) 535–538.

[40] G. Moretto: Manuale di audiologia. Libreria Editrice universitaria Levrotto & Bella, Turin, 1960.

[41] K. S. Pearsons, R. L. Bennett, S. Fidell: Speech levels in various noise environments. EPA-6001-77-025, US Environmental Protection Agency, 1977.

[42] I. R. Cushing, F. F. Li, T. J. Cox, K. Worrall, T. Jackson: Vocal effort levels in anechoic conditions. Applied Acoustics **72** (2011) 695–701.