PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Acoustic and subjective evaluation of Brazilian Portuguese speech recordings made in critical listening environments

Ticiana Matar DE LELLO[1]; Stephan PAUL[2]; Luiz Wagner Pereira BISCAINHO[3]

[1] PEE/COPPE - Federal University of Rio de Janeiro, Brazil

[2] Mechanical Engineering Department - Federal University of Santa Catarina, Brazil

[3] PEE/COPPE & DEL/POLI - Federal University of Rio de Janeiro, Brazil

ABSTRACT

The present work aims to characterize Brazilian Portuguese speech in realistic environments by means of acoustical parameters and subjective perception, hoping to provide data for research on auditory prostheses. Speech signals were recorded in several non-controlled environments such as restaurants, classrooms and stores. Sound pressure level, loudness and spectral content were extracted for speech and background noise separately by using recording segmentation and spectral subtraction. Subjective perception was analyzed by applying listening tests with samples from the recordings. Data analysis showed an increase of 0.6 dB in the speech's SPL per each dB increase in the background noise's SPL and a great difference in overall loudness between the quietest and the noisiest environments. SPLs were also mostly coherent to those previously reported in the literature for the same type of environment. The listening tests' results, while also greatly influenced by personal taste, showed that the annoyance caused by background noise, impairment in speech intelligibility, and the effort to understand speech were correlated to the background noise level and the SNR. In general, the environments perceived as the most critical for spoken communication were those in which background noise presents the highest energy concentration in the spectral region of speech.

Keywords: Speech recordings, Acoustic parameters, Subjective tests

INTRODUCTION

Auditory prostheses, such as hearing aids (HA), middle ear implants and cochlear implants (CI) are essential for hearing impaired individuals worldwide. They offer a better hearing experience in the user's everyday life and can provide an increase in life quality. However, even if auditory prostheses possibly help in increasing audibility and speech intelligibility in quiet environments, it can be seen that users usually still experience problems in environments with significant background noise and reverberation (1) or in situations with several speakers talking simultaneously.

As these characteristics are very frequent in daily life, it is necessary to study the acoustical characteristics of speech and background noise in everyday situations and its influence on subjective perception, in order to help developing prostheses which are more adapted to realistic acoustical scenarios. This is even more urgent for Brazilian prostheses' users, due to the lack of research on this topic considering Brazilian Portuguese speakers, Brazilian environments and the peculiarities of their cultural and socio-acoustical contexts.

In this work, recordings of speech signals were obtained in multiple non-controlled environments common to everyday life, such as classrooms, restaurants and commercial establishments. Speech and background noise levels were extracted from the recordings, analyzed and compared with previously obtained data (2, 3, 4). Loudness levels and spectral content for speech and background noise were also computed. Lastly, listening tests were performed using sound samples from the recordings to verify the association between subjective perception and acoustic parameters.

[1] ticiana@poli.ufrj.br
[2] stephan.paul@ufsc.br
[3] wagner@smt.ufrj.br

1. METHODS

1.1 Extraction of acoustic parameters

First, calibrated binaural speech signal recordings were obtained in seven restaurants (R1–R7), five classrooms (C1–C5), four commercial establishments inside a shopping mall (M1–M4), a living room in a domestic environment (LR1, LR2), an airport's boarding area (A1) and several environments within a university campus (O1–O3). The recordings were then divided into small segments where the main speaker's speech is present (regarded as "speech segments") and segments where it is absent and there is only background noise, composed by all the other sound sources in the acoustic scene, including other people's voices (regarded as "non-speech segments"). This allowed the extraction of acoustic parameters of speech and background noise separately. The segments were cut alternately (non-speech, speech, non-speech,...) and have durations which span from tens of milliseconds up to one minute. This was done manually with the HEAD Acoustics' Artemis SUITE® software.

Sound pressure level estimates were then obtained for speech and background noise for each recording, or part of recording. The calculations were done equally for A and C frequency weightings (5). An equivalent background noise level $L_{X\text{eq,noise,NS}}$ was obtained for the set of non-speech segments (NS), and an equivalent overall level $L_{X\text{eq,speech+noise,S}}$, which corresponds to the main speaker's speech and background noise combined, was obtained for the set of speech segments (S) in the recording. Letter X here is a placeholder for A- or C-weighting generically. Therefore, a background noise SPL estimate $\hat{L}_{X\text{eq,noise,S}}$ is calculated for a speech segment i based on the previous and subsequent non-speech segments, respectively, as follows:

$$\hat{L}_{X\text{eq,noise,S},i} = 10\log[1/2(10^{0.1L_{X\text{eq,noise,NS},i-1}} + 10^{0.1L_{X\text{eq,noise,NS},i+1}})] \tag{1}$$

A SPL can now be estimated for the sole speech contribution in this segment:

$$\hat{L}_{X\text{eq,speech,S},i} = 10\log[1/2(10^{0.1L_{X\text{eq,speech+noise,S},i}} - 10^{0.1\hat{L}_{X\text{eq,noise,S},i}})] \tag{2}$$

This procedure was done for each group of three segments in the sequence NS−S−NS (non-speech, speech and non-speech segment, respectively), regarded as triples, inside each recording, and for A and C weightings. In some cases, the energetic mean that corresponds to the background noise level estimate $\hat{L}_{X\text{eq,noise,S}}$ in the speech segment was found to be greater than the segment's total level $L_{X\text{eq,speech+noise,S}}$, when the latter is smaller than the total levels of the adjacent non-speech segments. These cases, which are less than 5% of all the segments, were considered invalid and removed from the sound pressure level analyses.

Overall Zwicker loudness over time for each recording was also calculated, which allowed to obtain an equivalent loudness for background noise for non-speech and speech segments. This was done with a similar method to that described in Equations 1 and 2, but using the arithmetic instead of the energetic means. The impossibility to isolate the loudness contribution of speech alone due to the complexity of the loudness model allowed only a comparison between the background noise loudness $\hat{N}_{\text{noise,S}}$ and the overall loudness $N_{\text{speech+noise,S}}$ of the speech segment, composed by background noise and the main speaker's speech.

Lastly, estimates of the speech and background noise spectra were obtained for each recording, with unweighted SPLs (5). An equivalent background noise spectrum for each recording was obtained as an energetic mean of the spectra of the recording's non-speech segments (which carried information of background noise only). The same procedure was repeated for the speech segments, generating an equivalent spectrum for the speech and background noise combination. This allowed an estimation for the recording's equivalent speech spectrum by using the spectral subtraction procedure described in Equation 3, where i is a spectrum bin, $L_{\text{speech+noise}}$ is the bin's total level, L_{n} is the bin's background noise SPL and L_{s} is the computed speech SPL. A factor $\beta = 0.01$ was chosen empirically to correct occasional results below 0 dB. The resulting spectrum is then trimmed to the fundamental frequency region (approximately 100 Hz for men and 200 Hz for women), removing frequency content below the fundamental frequency.

$$L_{\text{s},i} = \begin{cases} 10\log(|\beta \cdot 10^{0.1L_{\text{s+n},i}}|), & \text{if } 10^{0.1L_{\text{s+n},i}} - 10^{0.1L_{\text{n},i}} < 0; \\ 10\log(10^{0.1L_{\text{s+n},i}} - 10^{0.1L_{\text{n},i}}), & \text{otherwise.} \end{cases} \tag{3}$$

1.2 Listening tests

In order to analyze the relationship between acoustic parameters and subjective perception, a listening test was implemented with 23 audio samples from some of the previously obtained recordings. The listening test was completed by 22 students of the Federal University of Rio de Janeiro−eight women and fourteen men−aged between 22 and 29, who reported no hearing loss. The tests were conducted in an acoustically treated room. Each subject should listen to the audio sample, always played back over the same circum-aural headphones, with calibrated volume and preservation of the two original audio channels. Answers were collected by a computer interface.

Before starting the tests, all subjects read a manual containing all the procedure's instructions. The manual made explicit that there was always a main speaker in each sample, whose voice was in the foreground and who spoke most of the time, even if there were occasionally other speakers in the background. Each sample had at least one and no more than two main speakers. Subjects were therefore instructed to identify the main speaker in the sample and answer the items considering the main speaker's speech as reference. Voices from other speakers eventually present in the sample should be then considered background noise, along with the other sounds in the sample.

The test's interface allowed a sample to be played as many times as desired and displayed three items (questions):

1. What is the degree of annoyance of the ambient noise, disregarding the main speaker's speech?
2. To which degree is the intelligibility of the main speaker's speech hindered by the ambient noise?
3. How much effort is required to understand the content of the main speaker's speech?

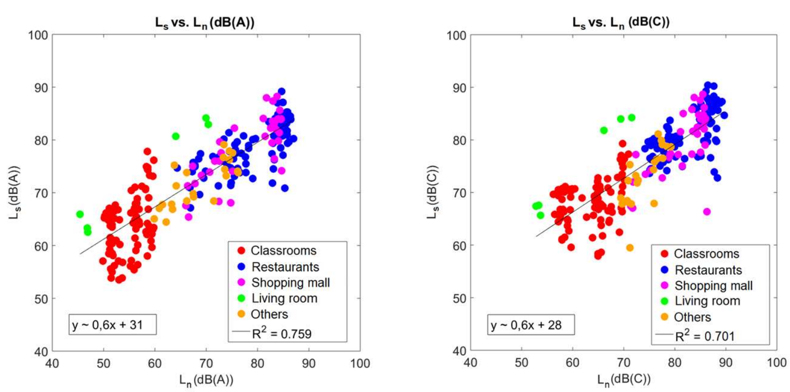
Each item was associated to a line scale whose ends were labeled NONE/NOTHING and EXTREME/EXTREMELY. The subject was instructed to place the cursor at the position in the scale which best represented his/her perception. The interface also displayed an "Observations" field, which could be optionally completed in case the subject desired to make comments or share any impressions regarding the sample.

The complete test featured twenty-three runs, each run with one audio sample of about 30 seconds. The first three runs were used for training and therefore removed from the results. The next twenty samples were then played in random order.

2. RESULTS AND DISCUSSION

2.1 Analysis of acoustic parameters

The A- and C-weighted speech SPL $\hat{L}_{X\mathrm{eq,speech},S}$ and noise SPL $\hat{L}_{X\mathrm{eq,noise},S}$ obtained by Equations 1 and 2 for each speech segment are plotted in Figure 1. An approximately linear relationship between the two variables can be seen for both weightings. Linear regressions were performed with the least-squares method, in which R^2 is the squared Pearson's linear correlation coefficient (6).



(a) A-weighting (b) C-weighting

Figure 1 – L_s versus L_n. Each environment type is represented by a particular color. Speech SPL as a function of noise SPL shows a growth rate of 0.6 dB/dB. $R^2 > 0.7$ for both weightings.

Both graphs show a 0.6 dB increase in speech level for each dB increase in noise level. This rate corresponds to that found in Pearsons *et. al.*'s study (2), in which A-weighted speech and noise SPLs were measured in many different environments. For background noise SPLs above 48 dB(A)−which is the case for all environments studied here−the authors observed this same rate of 0.6 dB/dB for all face-to-face communication situations in all environments with the exception of classrooms. This rate does not apply to each environment type in this work separately, but to the entire set of points that represent the noise and speech levels in all environments. Due to the similar relations found for both A- and C-weightings and to the comparison with other studies which use A-weighting only, it was decided to present only the A-weighted sound pressure levels for the subsequent sound pressure level analyses shown in this article.

Table 1 shows the arithmetic means $\bar{L}_{A\mathrm{eq,noise}}$ and $\bar{L}_{A\mathrm{eq,speech}}$ in some of the environments. The table also shows some speech and noise SPLs found in other studies for comparison: "Living room (casual talk)" and "Shopping mall (stores)" are compared to the levels obtained by Pearsons *et. al.* (2) for domestic environment and department stores, respectively; "Restaurants", to those of To and Chung (3); "Shopping mall (snack bar)", to an average of the noise levels measured in the food courts of four shopping malls by Carvalho and Pereira (4); and "Classrooms", to the averages of speech and noise levels found for two classrooms in Pearsons *et. al.* (2) for a speaker-listener distance of 7 meters. Restaurants are classified by occupational density, considered, according to To and Chung's criteria (3), as high (80-100% occupation), medium (60-79%) or low (40-59%). Speakers and listeners are at face-to-face communication distance in most situations (0.3−1 m), with the exception of the classroom scenario (5−9 m) and the living room scenario, in which the distance is 1 meter for the measurements in Pearsons *et. al*. (2), and within the range of 2 to 5 meters in the recordings performed in this study.

Table 1 – Average speech and background noise levels in different environments

Environment	$\bar{L}_{A\mathrm{eq,noise}}$/Other studies (dB(A))	$\bar{L}_{A\mathrm{eq,speech}}$/Other studies (dB(A))
Restaurants (high)	79/77 (3)	78/78.58 (3)
Restaurants (medium)	76/73 (3)	77/75.53 (3)
Restaurants (low)	68/69 (3)	74/72.83 (3)
Living room (casual talk)	46/48 (2)	64 (2-5 m)/57 (1 m) (2)
Shopping mall (stores)	70/54 (2)	74/61 (2)
Shopping mall (snack bar)	83/69 (4)	82
Classrooms	56.2/51 (2)	63.1/59.5 (2)

Substantial differences between levels measured in this and other studies can be seen for the shopping mall environments (stores and snack bar). For stores, in particular, a higher concentration of people and loud music in the background could be possible causes for the higher speech and noise SPLs in this work, in comparison to those present in that type of environment by the time of Pearsons et. al.'s work (2), in 1977. The noise SPL found in this study for the snack bar, on the other hand, is more similar to that found in highly occupied restaurants in To and Chung (3) and to that measured in food courts by Carvalho and Pereira (4). A possible cause for the higher speech SPL in the living room in comparison to that of Pearsons *et. al.* (2) was the raised vocal effort used by the speaker in this study, who was also further away from the listener.

Speech and noise SPLs in classrooms also have slightly greater values in comparison to those of Pearsons *et. al.* (2). The acoustical conditions of the classrooms in that study is unknown, but climate differences between United States and Brazil suggest that the classrooms used by Pearsons *et. al.* for measurements had no influence of air conditioner or fan noise, and probably feature better sound absorption and sound insulation, which also reduce the influence of external noise.

The relation between the background noise loudness and the total loudness is also similar to that found for speech and background noise SPLs. There is a growth rate of 1.1 sone/sone for the total loudness as a function of the background noise loudness. For the quietest places, such as classrooms, there is an average total loudness of 16 sone, against 9 sone from background noise only. In the noisiest restaurants, on the other hand, the total loudness can exceed 60 sone, while background noise loudness can be of 50 sone or even higher. Therefore, background noise in the noisiest environments of this study can be perceived as up to six times louder than that found in the quietest ones.

The speaker's contribution to the total loudness, however, decreases as the background noise loudness increases. One could understand that, in an already very noisy place, a small increase in the total sound energy produced by the speaker's voice would not make as much of a difference for the listener's sound volume sensation as it would in a very quiet place. For the quietest environments studied here, background noise loudness corresponds to 50% of the total loudness, while it is only 17% smaller than the total loudness in the noisiest environments.

The 1/3-octave spectra also illustrate the acoustical differences within the environments and situations, as can be seen in Figure 2, for six different situations: living room in a casual talk situation (LR1), living room in a party conversation situation (LR2), a teacher giving a lecture in classroom C1, a speaker talking to his listener across the table in restaurant R4, a face-to-face communication inside a shopping mall boutique (M1), and the voice transmitted through loudspeakers announcing flights inside a boarding area of an airport (A1). Quiet environments, such as classrooms and the living room in the casual talk situation, have in common a "flatter" background noise spectral shape, clearly distinguished from the speech's spectral shape; speech SPLs rise clearly above background noise SPLs in most, if not nearly all, 1/3-octave bands. Even so, a great individual difference between speakers could be seen for some classrooms, such as classroom C5, which had a teacher and a student making a presentation speaking at the same position in front of the class in different moments of the same lecture. As seen in this example, different talkers in the same classroom produced different shaped speech spectra and each had a different average speech level.

Background noise in the noisier environments, such as restaurants, shopping mall establishments and the boarding area of an airport, on the other hand, had a more complex spectral shape. It can be seen that background noise in those places exhibits a high energy concentration in the same spectral band as speech, between 100 and 2000 Hz, approximately. Speech is then largely masked by background noise, which explains the speaker's need to often raise his voice significantly in an attempt to be heard and understood (Lombard effect). Sound sources in those noisy environments which contribute to the increase in background noise are often music (in case of stores, restaurants and the home party), typical noises associated to that environment (e.g. the sound of dishes and cutlery in restaurants), and lastly, and mainly, other people's voices in the background. The simultaneous speech of multiple background talkers is the main factor for the energy accumulation of background noise in the spectral region of the main speaker's speech and for approximating the spectral shape of background noise to the spectral shape of speech itself. This can potentially cause confusion in the listener, worsen the communication conditions and speech intelligibility.

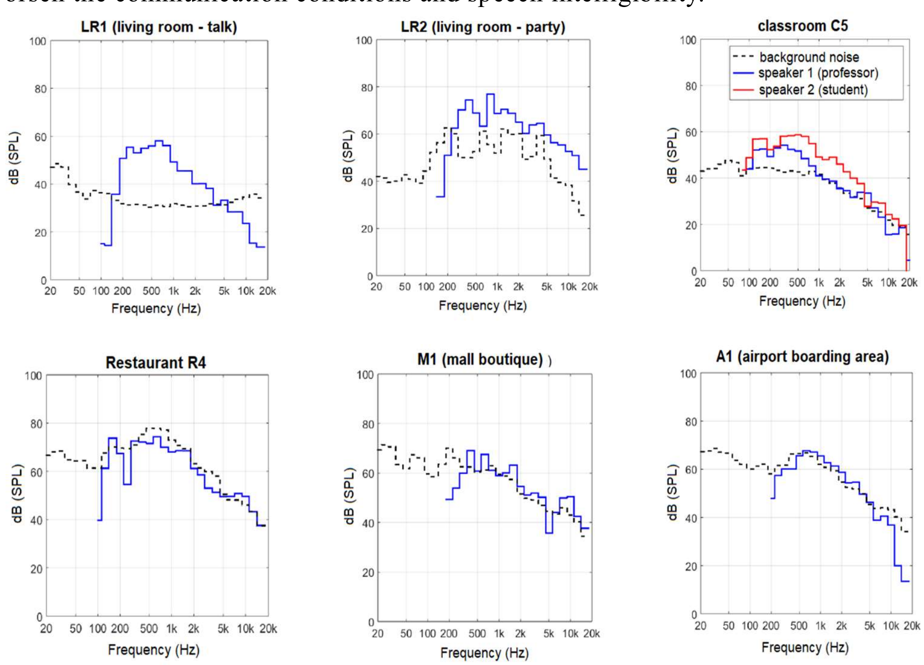


Figure 2 – Spectra of background noise (dotted line, black) and speech (solid line, blue or red) for some environments.

2.2 Subjective perception

The cursor positions chosen by the test subjects as answers for the three items in each test run were converted to integers between 0 (corresponding to the NOTHING end) and 100 (corresponding to the EXTREME/EXTREMELY end). The answers of each subject were analyzed. One of the criteria to consider a subject's response as an outlier was the presence of something that evidenced a lack of correct understanding of the test instructions. According to this, responses of two subjects were considered as outliers and removed from the data set, given that the incoherence in their answers seemed to show and their noticeable discrepancy with the other subjects' answers evidenced a difficulty in understanding the items' meanings and the line scales.

Problems with the identification of the main speaker occurred only for the sample A1 (airport), in which the main speaker actually corresponded to the voice announcing the flights through the loudspeakers, instead of any other speaker in the background, which confused some of the subjects. Due to this, only the answers of nine participants that had been explicitly warned about this detail before the test run were considered for this sample.

Figure 3 shows the average values and the standard deviation for the answers to each item. Classrooms C1, C2, C4 and C5 have two samples each. When there is the same main speaker in both samples, samples are distinguished by letters "a" and "b". When each sample has a different main speaker, a number is used (e.g. C1.1 and C1.2 refer to Speaker 1 and 2 in the same classroom at different moments). There is a large standard deviation for most items, showing the differences in individual perception. Even so, the average values behaved similarly to what was expected, with the most critical scenarios such as those in restaurants and at the snack bar inside the shopping mall (M4) receiving the worst (highest) grades in all items.

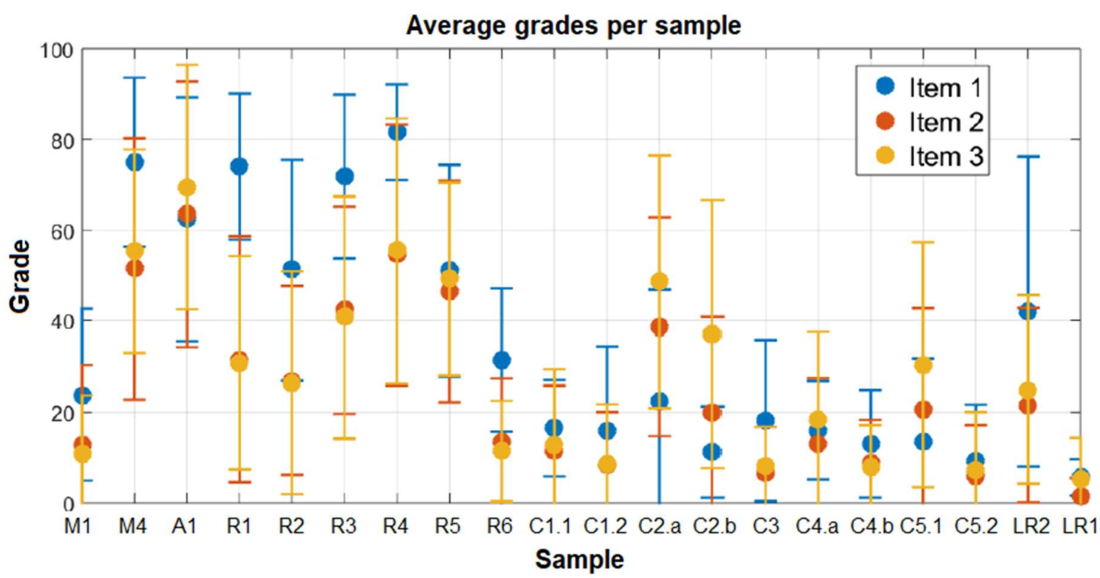


Figure 3 – Average punctuation per sample, for all three items.

There is strong correlation between the average grades of Items 2 and 3 ($r = 0{,}97$, $p < 0{,}05$), Items 1 and 2 ($r = 0{,}84$, $p < 0{,}05$), and Items 1 and 3 ($r = 0{,}72$, $p < 0{,}05$). As expected, a high degree of annoyance by the background noise is associated to greater impairment in speech intelligibility and greater effort to understand speech. On the other hand, a great amount of effort to understand speech is usually, but not always, associated to annoying background noise.

A demonstration that a great amount of effort to understand speech is not always related to background noise itself are the grades given to samples C5.1 and C5.2. Grades given to sample C5.2 are noticeably better (lower) due to the higher speech SPLs and better utterance of Speaker 2 (student), in comparison to Speaker 1 (teacher), even though background noise is low, not annoying and almost identical in both samples. Other cases which received moderately high grades for Item 3 despite receiving low grades for Item 1 have these grades possibly related to the speaker's utterance, speech speed, speech style and, in some cases, accent. For classrooms, a possible influence of reverberation time in speech intelligibility was also presumed.

On the other way around, grades given for Item 1 were higher than expected for cases which received moderately high grades for Items 2 and 3 – mainly, when the speech level is pretty low. This could mean that background noise may become more annoying simply by the fact that the listener cannot understand the speaker well enough.

Also very important was the influence of personal taste. Party noise in sample LR2 and the background music inside the boutique in sample M1, for example, were extremely annoying for some subjects, while others considered them actually not annoying at all. Some subjects reported impressions in the "Observations" field that also showed interesting aspects of subjective perception, such as annoyance caused by the cutlery noise in restaurants and by air conditioner noise in classrooms, sleepiness caused by softly-speaking teachers in classrooms and even having a hard time not concentrating on the music in the background in some samples.

Lastly, the grades given to each sample were compared to the previously obtained acoustical parameters and spectra of the corresponding recordings. Since the samples were chosen and edited in length for the listening tests after the acoustical parameter extraction from the recordings had already been made, and since a repetition of this procedure for the samples themselves could not be done for logistical reasons, the comparison was made with previously measured parts of the recordings which had the largest possible overlap with the samples created for the listening tests. Table 2 shows the correlation of the average grades received by the samples to the average speech and background noise SPLs, to the signal-to-noise ratio, to the average total loudness $\bar{N}_{\mathrm{eq,speech+noise}}$ and to the background noise loudness $\bar{N}_{\mathrm{eq,noise}}$, found for the parts of the recordings representing the samples.

Table 2 – Correlation of average grades given to Items 1, 2 and 3 and acoustical parameters

	$\bar{L}_{A,\mathrm{eq,speech}}$		$\bar{L}_{A,\mathrm{eq,noise}}$		SNR		$\bar{N}_{\mathrm{eq,speech+noise}}$		$\bar{N}_{\mathrm{eq,noise}}$	
	r	p	r	p	r	p	r	p	r	p
Item 1	0.82	<0.05	0.95	<0.05	-0.71	<0.05	0.95	<0.05	0.96	<0.05
Item 2	0.49	<0.05	0.74	<0.05	-0.77	<0.05	0.66	<0.05	0.75	<0.05
Item 3	0.3	0.2	0.59	<0.05	-0.74	<0.05	0.56	<0.05	0.61	<0.05

Item 1 is strongly correlated to all parameters, and mostly to those directly related to background noise itself. This shows that, indeed, the degree of annoyance caused by background noise is mostly proportional to the A-weighted background noise SPL. Nevertheless, it is important to note that the average sample duration of 30 seconds gives a very short time for the listener to accommodate himself to the ambient noise. With a longer accommodation time, grades given to the degree of annoyance caused by the background would be possibly lower. The listening test situation is also an artificial scenario, in which the subject focuses his attention to elements to which he otherwise would not in real situations. It must be considered once again that factors other than SPL and loudness can be related to annoyance, such as the timbre, the nature of the noise, and personal taste of the listener.

Item 2, while also strongly correlated to background noise SPL and loudness, is most strongly correlated to the SNR, and Item 3 is only strongly correlated to the SNR. These items, therefore, do not particularly relate to speech or background noise alone, but mainly to the way speech and background noise interact with each other–the more masked speech is by background noise, the more damaging background noise is to intelligibility, and the more effort is required by the listener to understand speech, even if background noise itself is not that loud after all.

A final comparison can be made between the grades and the average spectra of speech and background noise of the corresponding recordings, or parts of recordings representing the samples. The listening test grades tended to be always better (lower), especially for Items 2 and 3, for situations in which the background noise spectrum does not have as much energy concentration in the spectral region of speech or as much of a similar shape to that of the speech spectrum, as can be seen in Figure 2 for the casual talk situation in the living room (LR1) and for the conversation inside the boutique (M1), respectively.

In the noisiest environments, such the restaurants, the snack bar inside the shopping mall and the airport boarding area, the main factor for the bad (high) grades in all three items is the fact that background noise in those recordings is mainly composed by other people's speech. The great amount of information sharing a similar spectral content to that of the main speaker's speech masks the main speech considerably. As can be seen in Figure 2 for restaurant R4 and the airport boarding area (A1), the background noise spectrum in those cases is similar both in the 1/3-octave band levels and in shape to the main speech spectrum itself (higher levels between 100 and 2000 Hz, with a local maximum between 500 and 1000 Hz). As expected, this effect in fact contributed to confusion and discomfort felt by the listeners, represented by the subjects.

All the spectra obtained from the recordings and used for comparison with the listening test results can be found in De Lello (7), as well as the entire work description in more detail.

3. CONCLUSIONS

This paper aimed to characterize Brazilian Portuguese speech in real world environments by analyzing both the acoustical parameters and subjective perception regarding speech and background noise of binaural recordings obtained in several such environments.

Results showed mostly similarities in speech and background noise SPLs with data obtained by other researchers and reported in the literature. SPLs were also coherent to the loudness values.

Listening test results evidenced a tight relationship between the annoyance caused by background noise and its sound pressure level, while reduction in speech intelligibility caused by background noise and the effort to understand speech had a strong correlation to the SNR. Situations in which background noise has a high energy concentration in the spectral region of speech were perceived as the most critical for spoken communication.

Future works on this subject must consider a more detailed analysis of the social, behavioral and architectural factors in these spoken communication contexts, as well as a greater diversity of Brazilian Portuguese accents, given that this study was conducted in Rio de Janeiro and had mostly representatives of the local accent. Some tasks to be accomplished are extracting more information from the speech signal, relating it to vocal effort; identifying the most prominent frequency regions of speech in each type of environments; verifying the influence of diction, speech speed and accent in speech comprehension; and considering the influence of reverberation time in speech intelligibility.

The knowledge obtained on this subject could then contribute to the development of more adapted auditory prostheses for hearing impaired people in Brazil and abroad, capable of providing users a more comfortable and natural experience in everyday life communication situations.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001. Thanks also to the National Council for Scientific and Technological Development (CNPq), which gives support to the development of this work; to HEAD Acoustics®, whose equipment SQuadriga II® and software Artemis SUITE® allowed the recordings and the acoustical parameters' visualization; and to GROM Acústica & Vibração, for lending the equipment and software, helping making this work possible.

REFERENCES

1. Payton KL, Uchanski RM, Braida LD. Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. Journal of the Acoustical Society of America. 1994;95(3):1581-1592.
2. Pearsons KS, Bennett RL, Fidell S. Speech levels in various noise environments. Technical Report EPA-600/1-77-025, United States Environmental Protection Agency. 1977.
3. To WM, Chung, AWL. Restaurant noise: Levels and temporal characteristics. Noise & Vibration Worldwide. 2015;46(8):11-17.
4. Carvalho A, Pereira CFRT. Sound and noise in shopping malls. Proc NOISE-CON 16; 13-15 June 2016; Providence, USA 2016. p. 70-77.
5. International Eletrotechnical Comission. IEC Standard No 61672-1. Electroacoustics − Sound level meters − Part 1: Specifications[Internet]. Geneva: IEC; 2013 [cited 2019 April 19]. Available from: https://webstore.iec.ch/publication/5708
6. Deep R. Probability and Statistics. Oxford, UK: Academic Press; 2006.
7. De Lello TM. Caracterização da fala do português brasileiro em ambientes acústicos realísticos [B.Sc. Project]. [Rio de Janeiro (Brazil)]: Federal University of Rio de Janeiro; 2018.