Analysis of Factors Affecting the Auditory Attention of Non-native Speakers in e-Learning Environments

Nikesh Bajaj[1,2], Jesús Requena Carrión[1], Francesco Bellotti[2], Riccardo Berta[2] and Alessandro De Gloria[2]
[1]Queen Mary University of London, UK
[2]University of Genova, Italy
n.bajaj@qmul.ac.uk
j.requena@qmul.ac.uk
franz@elios.unige.it
berta@elios.unige.it
adg@elios.unige.it

Abstract: One of the most striking characteristics of e-Learning audiences is their diversity. Native and non-native learners can be expected among such audiences and therefore, when developing e-Learning courses it is important to consider the impact of the language level on learning. Specifically, non-native learners are expected to have a diminished auditory perception compared to native ones and hence reduced attention capabilities that could result in a poorer performance. In this study, we assess the impact of linguistic and auditory factors on the attention of non-native learners, namely semanticity, sentence length and noise level. An English language platform mimicking real e-Learning environments is used and attention is quantified by measuring the number of English words correctly identified during a listening task. Our results show that changes in each factor affect the attention score significantly. Interestingly, the effects of semanticity are apparent in noiseless environment, but vanish in noisy ones. Results also show that in noiseless environments, a change in the length of semantic sentences from small or medium to long causes a significant drop in the attention score. Our results demonstrate the importance of carefully accounting for linguistic and auditory factors when designing effective e-Learning courses, especially when they target global audiences and learners with different language abilities are expected.

Keywords: Auditory attention, e-Learning environment, non-native speakers, cognitive psychology, listening task

1. Introduction

The widespread use of multimedia contents and an increasing internationalisation of the classroom are two of the most prominent features of the current education landscape. Understanding their impact on learning can be critical to design effective learning experiences, especially in e-Learning environments where both features play such a central role. Multimedia contents and the internationalisation of the classroom have been a focus of interest in the education community and their impact on learning and student performance has been investigated empirically (Chang and Lehman, 2002; Guo, Kim and Rubin, 2014; Chen and Wu, 2015) and analysed based on cognitive theories (Moreno and Mayer, 1999; Cierniak, Scheiter, and Gerjets, 2009). From a cognitive perspective, multimedia approaches force learners to switch their attention between different sources and subsequently to carry out integration work. The cognitive strain associated to multimedia contents has led to a number of empirical studies that have investigated the impact of multimedia contents on learning (Fisher, Godwin, and Seltman, 2014; Aagaard, 2015) and to the formulation of guiding principles to be considered during the design of multimedia material, such as the split-attention principle (Ayres and Sweller, 2014) and the seductive detail principle (Rey, 2014). The role of the linguistic dimension in international learning environments has also awaken interest amongst the education community. Previous studies have focused on learning environments with native learners and non-native instructors (Nord, 1980; Evans and Alexander, 1984; Tseng, 1987). Adverse listening conditions such as accented speech and noise and their impact on intelligibility have been investigated by McLaughlin, et al. (2018) and the role of listeners' previous exposures to accented speech and the semantic context in the intelligibility of non-native speech has been studied by Kennedy and Trofimovich (2008). Language processing in non-native learners has also been explored. A study (Clahsen and Felser, 2006) concluded that even when non-native learners have an excellent grasp of the grammar, real-time processing of language, such as the one expected in a classroom environment, might still be a considerable challenge.

As with any other sensory modality, attention plays an important role when processing auditory information and therefore can have an impact on language processing and learning. For instance, it is well known that task-irrelevant auditory stimuli can undermine short-term memory and have a negative impact on language comprehension and learning (Hughes, et al., 2016; Marsh, et al., 2018). In addition, it has been suggested that

Reference this paper: Bajaj, N., et al., 2021. Analysis of Factors Affecting the Auditory Attention of Non-native Speakers in e-Learning Environments. *The Electronic Journal of e-Learning,* 19(3), pp. 159-169, available online at www.ejel.org

the mechanism of auditory selective attention might depend on the perceptual load (Murphy and Greene, 2017; Murphy, Spence, and Dalton, 2017).

E-Learning platforms are preferred over conventional ones by younger generations (Generation Z) (Rothman, 2016) and their use has increased as a result of the COVID-19 pandemic. As e-Learning is becoming widespread, it is vital to investigate on effective ways to design e-Learning environments. Recent studies suggest that Generation Z has a limited attention span and is easily distracted by visual stimuli (Shatto and Erwin, 2017; Poláková and Klímová, 2019). In contrast to other passive process, auditory perception is an adaptive process that continuously evolves to improve the identification of new sounds, a feature that is known as auditory learning (Heald, Van Hedger, and Nusbaum, 2017). Auditory learning is even more crucial for non-native learners that are interacting with e-Learning environment, as they are continuously exposed to new sounds from a second language. Auditory attention has been linked to learning and development in children (Gomes, et al., 2000) and considered as an important factor to understand complex acoustic scenes such as the multimedia ones used in e-Learning and classroom teaching environments (King, Teki, and Willmore., 2018). Thus, understanding auditory attention in non-native audiences is critical for designing effective e-Learning environments. However, the current literature shows a lack of quantitative studies focusing on the factors to be considered when designing learning environments or giving a lecture. In this article, we investigate how auditory attention of non-native speakers in an e-Learning environment is affected by linguistic and auditory factors. As part of our design of an e-Learning experience, we create a lab controlled experiment and identify potential factors affecting auditory attention, namely background noise, semanticity and length of auditory stimuli (Compton, 1967; Killion, 2002). We conduct the study with 25 healthy participants. Our study is relevant both to the education community, as we provide additional insight that can be used when designing learning experiences, and to the cognitive psychology community, as by focusing on non-native listeners our results extend previous findings in auditory attention.

The rest of the article is organized as follows. Section 2 reviews previous literature on auditory attention. Section 3 describes our experimental setting and Section 4 our experimental procedure. In Section 5, we present the statistical methods used for analysing our data. In Section 6 we analyse the results of our experiments and Section 7 provides our main conclusions.

2. Related Work on Auditory Attention

Auditory attention became a topic of great interest in cognitive and neuroscience circles in the years following the Second World War, aroused by the documentation of cases where audible messages failed to be perceived by fighter pilots (Spence and Santangelo, 2010). Most of the studies focusing on auditory attention involved simulated multi-message environments (Anderson, 1985) and among them, the dichotic listening task has become one of the most popular settings. In the dichotic listening task, two different auditory stimuli are presented to the participants, who are asked to attend to only one of them (Ingram, 2007). After listening to the stimuli, participants are asked to write down the messages that have been previously presented to them, and the written messages are used as an indication of the level of auditory attention (Cherry, 1953).

A major focus of research in auditory attention has been the understanding of selective attention, change deafness and spatial attention. Selective attention investigates why and when the brain selects and pays attention to one of the many stimuli available (Mayer, et al., 2006; Wu, et al., 2007) and is a concept that has also been successfully applied to other sensory systems, such as the visual (Shinn-Cunningham and Barbara, 2008; Dalton and Spence, 2008). Two main theories have been proposed for explaining the mechanism of attention. The first one is known as filter theory and suggests that the brain filters useful information over useless information (Broadbent, 2013). Two opposing mechanisms have been proposed in the context of filter theory, namely early selection theory and late selection theory (Treisman, 1964). According to the early selection theory, the brain selects stimuli at early stages of information processing, whereas late selection theory suggests that the brain selects stimuli after semantic decoding only, i.e. at a later stage of information processing (Deutsch and Deutsch., 1963). The second theory which provides a model for auditory attention is load theory and it gives a plausible unifying framework for early and late selection theory (Lavie, 2005). According to load theory, selection can be at early or late stages, depending on the perceptual load of the stimuli (Lavie, 2005). Specifically, if the perceptual load of the stimuli is high, the brain will filter out useless information at early stages, whereas if the perceptual load of stimulus is low, then all the stimuli are processed before being filtered out, which corresponds to a late selection process (Deutsch and Deutsch., 1963).

Auditory change deafness refers to the mechanisms by which the brain misses information in auditory stimuli during short time intervals. Auditory change deafness is analogous to the phenomenon of psychogenic blackout or transient loss of consciousness (Fitzpatrick and Cooper, 2006) in that there is a short interval in continuous listening processes, during which the brain does not process any auditory information. Previous auditory attention studies have focused on the effect of the length of stimulus and the complexity of sentence (Compton, 1967) and background noise (Killion, 2002). The authors in (Compton, 1967) demonstrated that simple structural English sentences were more easily perceived than complex structural ones, however the length of the sentence did not have a considerable impact on it. In (Killion, 2002) it was concluded that the effect of noise on perception of speech diminishes with increased signal to noise ratio (SNR). Teraoka, et al. (2018) investigated how the spatial origin of auditory signals affect word intelligibility in complex acoustic scenes. In (Bai, Zhao, and Xie, 2019), the subjective auditory attention saliency of 20 pieces of sound material was evaluated in a psychoacoustic experiment. Their results show the importance of adapting computational models based on the characteristics of the targeted sound types. Another important aspect for education concerns joint attention, which consists of two people simultaneously and jointly focusing on an item. While our work does not touch this point, we acknowledge the key importance of human relations for learning. In the present study we extend previous experimental results by investigating auditory attention in a group of non-native listeners who are subjected to different auditory conditions in a simulated e-Learning environment.

3. Methods and materials

3.1 Participants

A group of 25 (21 male, 4 female) healthy, university students of science and technology, with no known auditory processing disorder were chosen for the study. The participants came from different nationalities (see Table 1) and their first language was different from English (see Table 2). The age distribution of the group of participants is shown in table 3.

Table 1: Nationality of the participants

Nationality	Number of participants
Algerian	1
Indian	8
Iranian	3
Italian	4
Kazakh	1
Lebanese	4
Moroccan	1
Nepalese	1
Pakistani	1
Tunisian	1

Table 2: First language of the participants

First language	Number of participants
Arabic	7
Farsi	3
Italian	4
Kannada	1
Kazakh	1
Mathili	1
Malayalam	4
Marathi	1
Tamil	1
Telgu	1
Urdu	1

Table 3: Age groups

Age Group (years)	Number of participants
16-20	1
21-25	6
26-30	16
31-35	2

3.2 Audio stimuli

A total of 5000 audio clips were obtained from the Tatoeba Project (Trang Ho, 2006) along with their corresponding text. All audio clips were English sentences of length ranging from 3 words to 13 words per sentence, one sentence per audio clip, and used the same male voice to ensure consistency in the physical properties of the audio signal (e.g. pitch, rate of speech). From this collection of audio clips, 1700 non-semantic audio clips were generated by suitably inserting isolated words in the original, semantic sentences. We ensured that the length of non-semantic sentences also ranged from 3 words to 13 words. The following sentences S1, S2, S3, and S4 are examples of semantic and non-semantic sentences:

S1: *She looks unhappy*.
S2: *You should have left half an hour earlier.*
S3: *Let's <u>intelligent</u> <u>next</u> go.*
S4: *I can <u>hour</u> ski <u>so</u> <u>late</u>.*

Sentences S1 and S2 are semantic whereas S3 and S4 are non-semantic. The generated non-semantic sentences are partially semantic, as they are made from semantic sentences. For example; sentence S4 is made by inserting the words hour, so and late to the semantic sentence I can ski.

We used this approach for generating non-semantic sentences rather than simply listing unrelated words as in (Baddeley, 1966), to mimic a real-world scenario in learning environment, where a few words that are unknown to the listener (out of vocabulary of listener) might make a sentence meaningless. This situation is commonly experienced by non-native speakers due to a limited vocabulary. In learning environments, a reduced subject-specific vocabulary can make semantic sentences be perceived by students as non-semantic. In addition, our approach for constructing non-semantic sentences can also reproduce other prominent features of live speech, notably stuttering.

Finally, background noise of different levels was also added to both semantic and non-semantic audio clip groups. SNR values of -6 dB, -3 dB, 0 dB, 3 dB, 6 dB and ∞ dB (noise-free case) were considered. In summary, we produced two groups of audio stimuli, i.e. the semantic and non-semantic groups, which were reproduced with six levels of background noise.

3.3 Length of sentence

Sentences were grouped according to their length in three categories, namely small (L1), medium (L2) and long (L3) sentences with average lengths of 4, 8 and 12 words respectively with variation of ±1 word. In this, we followed closely the sentence lengths used in (Compton, 1967) for investigating aural perception. In allowing different sentence lengths within each group, we recognized that some words are phonetically longer than others, i.e. 'I' and 'Congratulation', and assumed that a difference of one word in sentence length does not have any significant impact on the listening task (Neath and Nairne, 1995; Page and Norris, 1998).

3.4 Case generation

A total of 144 English sentences were generated for each participant by extracting 72 sentences from the pool of 5000 semantic sentences and 72 more sentences from the pool of 1700 non-semantic sentences. Specifically, the semantic (resp. non-semantic) group consisted of 30, 24, and 18 sentences of lengths L1, L2 and L3 that were extracted randomly without replacement from the pool of semantic (resp. non-semantic) sentences. Each length sub-group was further subdivided into six equal-sized sub-groups, each one of which was assigned one of the six noise levels described in Section 3.2. Each sub-group was ensured to have the respective average length of sentence, for instance, the average length of L1 sub-groups is 4 words. This resulted in a total of 36 experimental conditions, corresponding to six levels of noise, three sentence lengths, and two semanticity levels ($6 \times 3 \times 2 = 36$). The number of audio stimuli per experimental condition was five for the L1 group ($30/6 = 5$), four for the L2 group ($24/6 = 4$) and three for the L3 group ($18/6 = 3$). This description is summarized in Table 4.

For readability purposes, each experimental condition is labeled as 'xxdByLz', where xxdB indicates the noise level, y is a binary digit 0 or 1 indicating respectively semanticity or non-semanticity, and Lz corresponds to the length of the sentence, namely L1, L2 or L3. For example, the label -6dB0L1 indicates the condition with -6dB SNR for semantic sentences of length L1.

Table 4: Number of sentences per experimental condition

	Semantic			**Non-Semantic**		
SNR	**L1**	**L2**	**L3**	**L1**	**L2**	**L3**
-6 dB	5	4	3	5	4	3
-3 dB	5	4	3	5	4	3
0 dB	5	4	3	5	4	3
3 dB	5	4	3	5	4	3
6 dB	5	4	3	5	4	3
∞ dB	5	4	3	5	4	3
Subtotal	30	24	18	30	24	18
Total	72			72		

4. Experiment procedure

Participants were presented with a computer interface and were given a passive headphone set. After entering basic demographic information, namely sex, nationality, age-group and first language, participants could initiate the listening task by clicking a play button. During the listening task, the computer interface remained disabled so as to prevent participants from engaging in other activities. Once the audio file had finished no more reproductions were permitted and participants were allowed to submit a transcription of the audio sentence. Upon submission, participants could reproduce the next audio file. This procedure was carefully explained to each participant.

On the average, the procedure involving the 144 audio files described in Section 3.4 took 35 minutes for a participant. At the end of each experiment, text transcriptions were collected and labelled according to the noise level, the length of the original English sentence and its semanticity.

5. Statistical analysis

5.1 Attention score computation

Based on an original English sentence and the text transcription of its corresponding audio, the attention score of the text transcription was calculated as follows. Let $T_{i,p,k}$ denote the $i-th$ transcription produced by the $p-th$ participant under the $k-th$ experimental condition. Then, its attention score, which we denote by $A_{i,p,k}$ was calculated as

$$A_{i,p,k} = \frac{N_{C(i,p,k)}}{N_{T(i,p,k)}} \times 100, \tag{1}$$

where $N_{C(i,p,k)}$ is number of correct words in $T_{i,p,k}$ and $N_{T(i,p,k)}$ is number of total words in the original sentence. When establishing the number of correct words in a transcription, minor errors in spelling and other typos were ignored, for example *looks/look*, *beautiful/beutiful* or *designed/disegned*.

Based on the attention score of each individual transcription, the average attention score $A_{p,k}$ of the transcriptions produced by the $p-th$ participant under the $k-th$ experimental condition was calculated as

$$A_{p,k} = \frac{1}{I_k}\sum_{i=1}^{I_k} A_{i,p,k}\,, \tag{2}$$

where I_k is the total number of stimuli in $k-th$ experimental condition (see Table 4). Each computed average attention score $A_{p,k}$ was used as an individual sample for statistical analysis. As $N_P = 25$ participants were involved in the experiment and each participant produced a total of $N_K = 36$ attention score samples (one sample per experimental condition), a total of 900 samples $(36 \times 25 = 900)$ were available for further analysis.

5.2 Descriptive statistics

Firstly, the mean A_k and standard deviation S_k of the attention score $A_{p,k}$ across all the participants were calculated:

$$A_k = \frac{1}{N_p}\sum_{p=1}^{N_p} A_{p.k}\,, \tag{3}$$

$$S_k = \left(\frac{1}{N_p - 1}\sum_{p=1}^{N_p}\left(A_{p,k} - A_k\right)^2\right)^{\frac{1}{2}}. \tag{4}$$

Secondly, the mean attention score A_p for the $p - th$ participant across all the conditions was computed. The mean attention score A_p was defined as

$$A_p = \frac{1}{N_k}\sum_{k=1}^{N_k} A_{p.k}\,. \tag{5}$$

Finally, box-and-whisker plots were obtained for analysing the impact of the noise level, the length of stimuli, and sentence semanticity on the attention score of the transcriptions from each experimental condition.

5.3 Pairwise Analysis

A repeated measure ANOVA test was firstly applied to analyse the variation of attention score under each experimental condition (Lawal, 2014). The student t-test was used to determine whether the mean of the attention score was significantly different under two different experimental conditions. Since, there are 36 different conditions, the p-values resulting from comparing each pair of experimental conditions were represented in a 36×36 P-matrix. By definition, this P-matrix is symmetric, and its diagonal represents the comparison of an experimental group with itself, hence the values in the diagonal are 1. In order to facilitate the analysis of the p-values, heat maps were used to visually represent the values of the entries of the P-matrix. Thresholding with a value of 0.05 was next applied to the P-matrix, producing a binary matrix in which significant differences with a 95% confidence level can be readily identified. The binary P-matrix allowed us to further arrange the experimental conditions in a hierarchical manner, following a bottom-up agglomerative method (Hastie, Tibshirani, and Friedman, 2001), where related experimental conditions are located close to one another.

6. Results and discussions

The mean A_k and standard deviation S_k of the attention score, across all the participants under each experimental condition are presented in Table 5. As expected, our results show that the lowest attention score was achieved for low SNR, whereas the high SNR was associated to high levels of attention. In general, an increase in the noise level produces a decrease in the attention score. This suggests a negative correlation between noise and attention, indicating the noise as a primary factor of diminished attention level. It is interesting to note that an increase in the length of a sentence resulted in a decrease in the attention score for any given noise level in both semantic and non-semantic groups. This suggests that the shorter sentences are, the easier to follow by listeners.

Table 5: Mean and standard deviation of the average attention score $A_{p,k}$ in each experimental condition

SNR	Mean ($n = 25$)						SD ($n = 25$)					
	Semantic			Non-Semantic			Semantic			Non-Semantic		
	L1	L2	L3	L1	L2	L3	L1	L2	L3	L1	L2	L3
-6 dB	13.03	5.86	4.85	10.41	8.21	6.91	14.03	6.53	6.74	9.74	6.51	7.10
-3 dB	33.41	21.12	15.56	21.57	15.13	11.61	23.02	17.02	8.21	14.96	10.04	7.40
0 dB	40.49	37.11	24.63	30.48	28.81	16.32	20.27	22.63	20.09	15	18.06	12.10
3 dB	57.09	49.17	43.24	38.91	32.9	22.38	24.88	26.21	25.74	18.17	19.19	14.54
6 dB	72.04	62.82	48.82	50.22	40.15	26.75	21.34	23.28	24.8	20.25	18.36	15.86
∞ dB	85.17	86.48	72.8	67.03	56.2	39.45	16.1	17.25	22.48	17.78	18.82	18.77

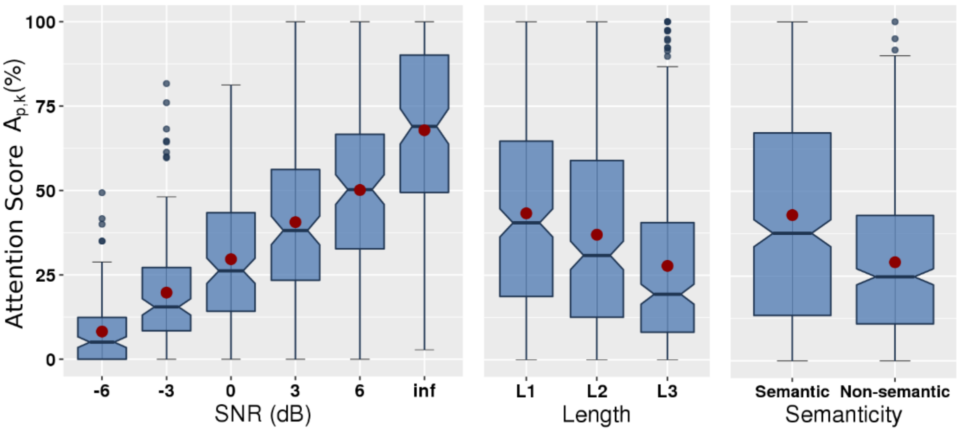
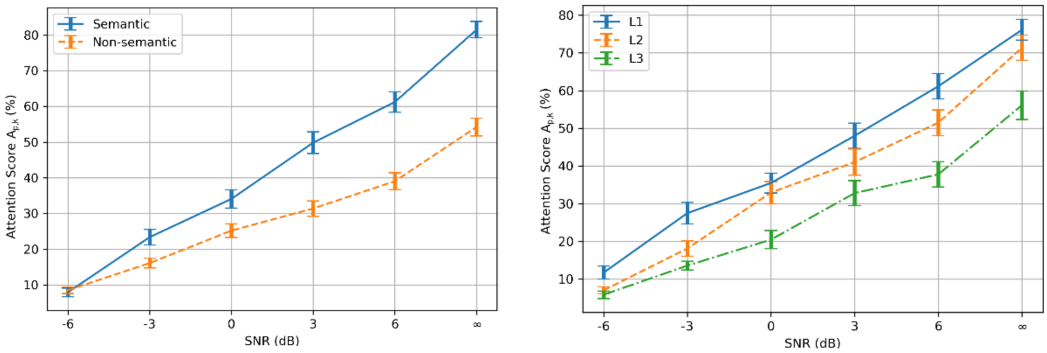


Figure 1: Average attention score $A_{p,k}$ versus SNR, length and semanticity of stimulus



(a) Noise and semanticity of stimuli

(b) Noise and length of stimuli

Figure 2: Interaction between independent variables

The effects of noise level and length of sentence on the attention score are shown in the box-and-whisker plots presented in Figure 1, demonstrating a clear dependence of the attention score on the SNR, sentence length, and semanticity. It is interesting to observe that in addition to noise level and semanticity, the length of sentence also affects the attention score.

The impact of semanticity and length of sentence on attention score, under different noise level was investigated, and results are shown in Figure 2. Interestingly, Figure 2 (a) indicates that the rate of increase of the attention score with increases in SNR is higher for the semantic group than for the non-semantic group. In other words, background noise has higher impact on semantic sentences than non-semantic. As a result, the effect of semanticity is more apparent in a noise-free environment and vanishes in a noisy environment. This effect can be explained by the cognitive neuroscience and linguistics Gap-Filling theory (Frazier, Godwin, and Seltman, 1983), according to which the brain is capable of predicting linguistic gaps by using syntactic or semantic priors (Tanenhaus, 1985; Frazier and d'Arcais, 1989). In other words, when the brain does not perceive a sentence correctly, as is the case in noisy environment, prior knowledge is used to infer it and fill the gaps. However, in noise-free environment, the brain is not forced to fill the gaps. The above observations suggest that the importance of semanticity is higher in noise-free environments. The non-overlapping error bars in Figure 2 (a) indicate that semanticity has a significant impact ($p < 0.05$) on attention score from -3 dB onward.

From Figure 2 (b), it can be concluded that the rate of change of the attention score for all the lengths is almost similar and smaller lengths produce consistently higher average attention scores. It suggests the length of a sentence always has an impact on the attention score for non-native speakers, irrespective to noise level. For native speakers, the length of sentence might not have such a significant impact.

Table 6: Results of the repeated measure ANOVA, where df denotes the degrees of freedom, SSq is sum of the squared differences, MSq is the mean sum of squares

Source	df	SSq	MSq	F-value	p-value
Between	35	456255.1	13035.86	72.77	10^{-16}
Subject	24	121625.8	5067.74	28.29	
Within	864	272098.9	314.93		
Error	840	150473	179.14		
Total	889	728354			

The results from our descriptive analysis suggest that the attention score is affected by noise level, semanticity, and length of sentence. The results from the repeated-measure ANOVA (Table 6), carried out for 36 experimental groups, indicate a strong evidence that there is a significant ($p \ll 0.001$) difference between at least two experimental groups. The resulting $\eta^2_{partial} = 0.752$, indicates that 75.2% differences in the experimental groups were due to different experimental conditions. In other words, the variation in the independent factors is largely responsible for the differences in the attention score.

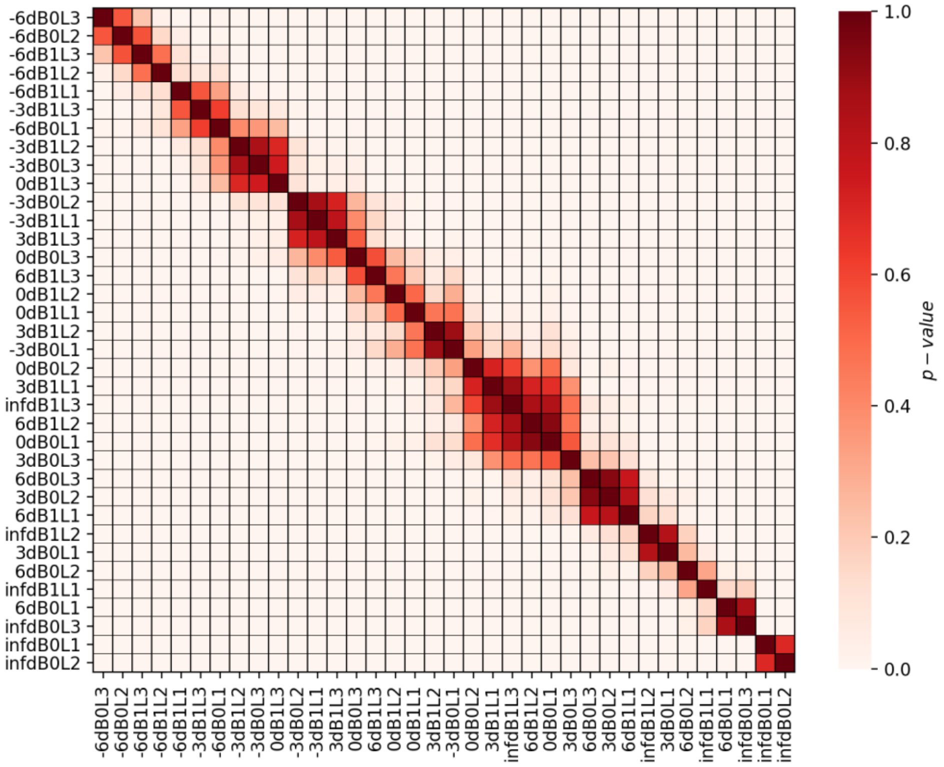


Figure 3: P-Matrix obtained from the pairwise t-test

Figure 3 shows the obtained P-matrix (36×36), sorted by the average attention score A_k of each experimental condition. As previously mentioned, by its definition the P-matrix is symmetric, and the values of the diagonal are identical to 1. In Figure 3, the thickness of diagonal represents the closeness of one experimental condition to other conditions. The thicker the diagonal, the more experimental conditions are closer to each-other. In an ideal situation where all the experimental conditions are significantly different from one another, the matrix would be close to a diagonal one. From Figure 3, it is apparent that a major fraction of pairwise comparisons have a low p-value, which suggests a significant difference between those respective pairs.

From Figure 3, clusters of experimental conditions can be identified by analysing the thickness of the diagonal. To analyse these clusters, a hierarchical clustering method is applied on the binary P-matrix obtained from thresholding with $p < 0.05$. Figure 4 illustrates the hierarchical tree obtained.

Given a branching point in the hierarchical tree, a cluster is defined as the collection of all the experimental conditions below the branching point. For instance, the branching point *R* in Figure 4 defines a cluster consisting of two groups, namely *infdB0L1* and *infdB0L2*, and both groups are close in the sense that the impact on the attention score by changing the experimental conditions from *infdB0L1* to *infdB0L2* is smaller than changing to an experimental condition outside the cluster *R*. The analysis of the hierarchical tree reveals interesting relationships about the experimental conditions in this study. For example, cluster *R* suggests that changing the length of semantic sentences in noiseless environment from small to medium or vice-versa has no significant impact on the attention level, however changing length to long sentence, switching semanticity or increasing noise level drop the attention level significantly ($p < 0.05$). This can also be confirmed from Figure 3, by looking at the rightmost bottom cluster of 2×2 cells.

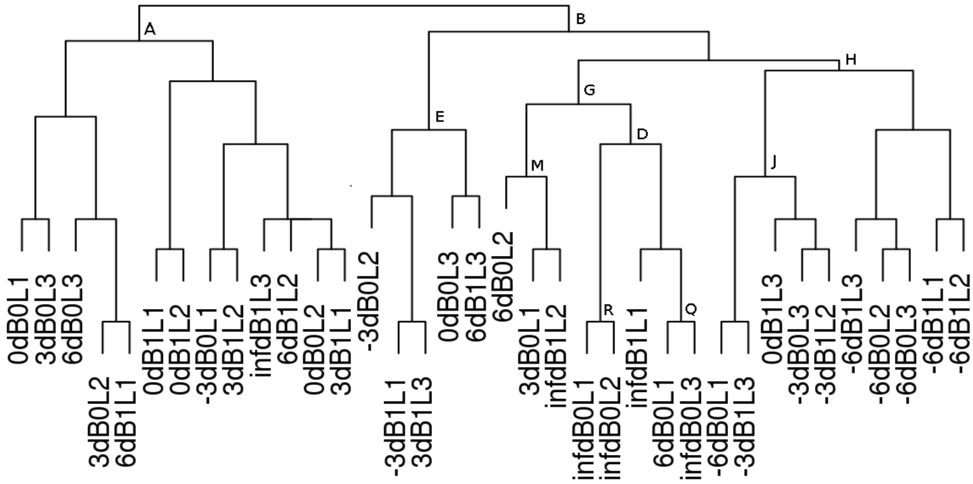


Figure 4: Hierarchically clustering of experimental conditions obtained after applying threshold $p < 0.05$ on P-Matrix

Furthermore, cluster Q (6dB0L1, infdB0L3) shows that long semantic sentences in noiseless environment have a similar impact on attention as small semantic sentences in low-noise (6 dB) environment. Similarly, cluster M suggests that the impact of small semantic sentences with SNR = 3 dB, is similar to medium-length non-semantic sentences in noiseless environments and semantic sentences in low-noise (6 dB) environments.

Cluster G includes the experimental conditions that lead to the highest performance. The mean attention score (A_k) of these groups is above 55. By contrast, cluster H includes the experimental conditions with the lowest mean attention score ($A_k < 17$). Interestingly, clusters H and A appear far from each other in the hierarchy and mostly differ in the noise level. Cluster H includes almost all the experimental conditions with low SNR, whereas cluster A includes experimental groups with high SNR and their mean attention score ranges from 27 to 51. Overall, the hierarchical tree can be decomposed into four major clusters, A, E, G, and H.

7. Conclusions

Auditory attention plays a vital role in students' performance and it is critical for non-native learners. The increasing number of international students require that education professionals carefully reflect on the design of their learning environments and their suitability for such an audience. However, the literature lacks quantitative results in the area. In this study, we have analysed the impact of three factors (noise level, length of sentences and semanticity) on the level of auditory attention of non-native learners in an English language environment. Following existing literature, we have used the number of correctly identified words in a transcribed sentence to quantify the level of auditory attention.

Our analysis emphasises the importance of conveniently designing lectures and learning environments and shows that linguistic and auditory factors such as background noise, length of sentences and lexical complexity should be carefully considered. In addition to the importance of an obvious factor such as background noise, our results demonstrate that reducing the length of sentences always improves the level of auditory attention of non-native speakers. The impact of linguistic complexity on the auditory attention of non-native learners can be analysed by considering sentence semanticity. In our experiment, non-semantic sentences were created by inserting unrelated words in-between semantic sentences. This model of non-semantic sentence simulates a semantic sentence that contains one or more words that are unknown to the listener, which is close to real-world scenarios. Our results show that semanticity has a dramatic impact on auditory attention of non-native learners and the significance of semanticity is higher in less noisy environment. This experimental observation supports the view that simple, accessible vocabulary should be consistently used in non-native learning environments, whenever this does not compromise the meaning of the message. Our results suggest that in practice, while delivering a lecture, shorter sentences should be preferred over longer ones. To keep the sentences meaningful to an audience, simplified and easily understandable (or well explained) terms should be preferred over new and complex terminology, whenever possible, and noise level should be controlled and kept as minimum as possible. Our analysis provides novel quantitative indications that should be useful to researchers, practitioners and designers for maximizing a students' experience in a learning environment. We focused our study on auditory attention, because it is fundamental for learning. Future research may expand on our results by including visual spatial attention and the relevant relationships that originate in real environments. The impact on lecture performance and instructional environment design can also be further investigated. Finally, future research work may analyse the possible differences in impact among students with different language abilities.

References

Aagaard, J., 2015. Media multitasking, attention, and distraction: a critical discussion. *Phenomenology and the Cognitive Sciences,* 14(1), pp. 885–896.

Anderson, J. R., 1985, Attention and performance. In: C. Woods, 2015. *Cognitive psychology and its implications.* 8th ed.New York: Worth Publishers. pp. 53-77.

Ayres, P. and Sweller, J., 2005. The split-attention principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning,* 2(1), pp. 135–146.

Baddeley, A. D., 1966. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology,* 18(1), pp. 362–365.

Bai, X., Zhao, Z., and Xie, L., 2019. Experimental analysis on auditory attention saliency calculation models*.* In: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), China, IEEE, pp. 1-5.

Broadbent, D. E., 2013. *Perception and Communication,* Kent UK*:* Elsevier.

Chang, M. M. and Lehman, J. D., 2002. Learning foreign language through an interactive multimedia program: An experimental study on the effects of the relevance component of the ARCS model. *CALICO journal,* pp. 81–98.

Chen, C.-M. and Wu, C. H., 2015. Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education,* 80(1), pp. 108–121.

Cherry, E. C., 1953. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America,* 25(5), pp. 975-979.

Cierniak, G., Scheiter, K., and Gerjets, P., 2009. Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?. *Computers in Human Behavior,* 25(1), pp. 315–324.

Clahsen, H. and Felser, C., 2006. How native-like is non-native language processing? *Trends in Cognitive Sciences,* 10(1), pp. 564–570.

Compton, A. J., 1967. Aural perception of different syntactic structures and lengths. *Language and Speech,* 10(1), pp. 81–87.

Dalton, P. and Spence, C., 2008. Selective attention in vision, audition, and touch. *Learning and memory: a comprehensive reference.* Amsterdam: Elsevier*;* pp. 243-57.

Deutsch, J. A. and Deutsch, D., 1963. Attention: Some theoretical considerations. *Psychological Review,* 70(1), pp. 80-90.

Evans, E. E. and Alexander, C. K., 1984. Increasing aural skills of international graduate students. *IEEE Transactions on Education,* 27(3), pp. 167-170.

Fisher, A. V., Godwin, K. E., and Seltman, H., 2014. Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science,* 25(7), pp. 1362–1370.

Fitzpatrick, A. P. and Cooper, P., 2006. Diagnosis and management of patients with blackouts. *Heart,* 92(4), pp. 559-568.

Frazier, L., Clifton, C., and Randall, J., 1983. Filling gaps: Decision principles and structure in sentence comprehension. *Cognition,* 13(2), pp. 187–222.

Frazier, L. and d'Arcais, G. B. F., 1989. Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language,* 28(3), pp. 331–344.

Gomes, H., Molhome. S., Christodoulou, C., Ritter, W., and Cowan, N., 2000. The development of auditory attention in children. *Frontiers in Bioscience,* 5(1), pp. 108-120.

Guo, P. J., Kim, J., and Rubin, R., 2014. How video production affects student engagement: An empirical study of MOOC videos. In: Proceedings of the first ACM conference on Learning, scale conference, pp. 41–50, New York.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The elements of statistical learning: Data mining, inference, and prediction* . New York: Springer Science & Business Media.

Heald, S. L., Van Hedger, S. C., and Nusbaum, H. C., 2017. Understanding sound: Auditory skill acquisition. *Psychology of Learning and Motivation,* 67(1), pp. 53-93.

Hughes, R. W., Chamberland, C., Tremblay, S., and Jones, D. M., 2016. Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory and Language,* 90(1), pp. 126–146.

Ingram, J. C. L., 2007. *Neurolinguistics: An introduction to spoken language processing and its disorders.* Cambridge*:* Cambridge University Press.

Kennedy, S. and Trofimovich, P., 2008. Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review,* 64(3), pp. 459–489.

Killion, M. C., 2002. New thinking on hearing in noise: A generalized articulation index., *Seminars in Hearing*, 23(1), pp. 47-62.

King, A. J., Teki, S., and Willmore, B. D., 2018. Recent advances in understanding the auditory cortex. *F1000Research,* 7(1).

Lavie, N., 2005. Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences,* 9(2), pp. 75-82.

Lawal, B., 2014. *Applied statistical methods in agriculture, health and life Sciences.* Cham: Springer.

Lawal, B., 2014. Repeated measures design. In: B. Lawal (ed.) *Applied statistical methods in agriculture, health and life sciences.* Cham: Springer International Publishing, pp. 697-718.

Marsh, J.E., Yang, J., Qualter, P., Richardson, C., Perham, N., Vachon, F., and Hughes, R.W., 2018. Post-categorical auditory distraction in short-term memory: Insights from increased task load and task type. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), p.882.

Mayer, A. R., Harrington, D., Adair, J. C., and Lee, R., 2006. The neural networks underlying endogenous auditory covert orienting and reorienting. *Neuroimage,* 30(3), pp. 938-949.

McLaughlin, D.J., Baese-Berk, M.M., Bent, T., Borrie, S.A., and Van Engen, K.J., 2018. Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, & Psychophysics*, *80*(6), pp.1559-1570.

Moreno, R. and Mayer, R. E., 1999. Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology,* 91(2), pp. 358.

Murphy, G. and Greene, C. M., 2017. Load theory behind the wheel; perceptual and cognitive load effects. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale,* 71(3), pp. 191.

Murphy, S., Spence, C., and Dalton, P., 2017. Auditory perceptual load: A review. *Hearing Research,* 352(1), pp. 40-48.

Neath, I. and Nairne, J. S., 1995. Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review,* 2(4), pp. 429–441.

Nord, J. R., 1980. Developing listening fluency before speaking: an alternative paradigm. *System,* 8(1), pp. 1–22.

Page, M. and Norris, D., 1998. The primacy model: A new model of immediate serial recall. *Psychological Review,* 105(4), pp. 761.

Poláková, P. and Klímová, B., 2019. Mobile technology and Generation Z in the English language classroom - A preliminary study. *Education Sciences,* 9(3), pp. 203.

Rey, G. D., 2014. Seductive details and attention distraction–An eye tracker experiment. *Computers in Human Behavior,* 32(1), pp. 133–144.

Rothman, D., 2016. *A Tsunami of learners called Generation Z.* Available at: <http://www. mdle.net/JoumaFA_Tsunami_of_Learners_Called_Generation_Z. pdf> [Accessed 24 April 2021].

Shatto, B. and Erwin, K., 2017. Teaching Millennials and Generation Z: Bridging the generational divide. *Creative Nursing,* 23(1), pp. 24--28.

Shinn-Cunningham, B. G., 2008. Object-based auditory and visual attention. *Trends in ognitive Sciences,* 12(5), pp. 182-186.

Spence, C. and Santangelo, V., 2010. Auditory attention. *Oxford Handbook of Auditory Science: Hearing,* 3(1), pp. 249.

Tanenhaus, M. K., 1985. The interaction of lexical expectation and pragmatics in parsing filler-gap constructions. In: Proceedings of the seventh annual meeting of Cognitive Neuroscience Association 1985, pp. 25–27.

Teraoka, R., Shuichi, S. Z. C., Suzuki, Y., and Shioiri, S., 2018. Influence of auditory selective attention on word intelligibility. In: 2018 Joint Conference-Acoustics., IEEE, pp. 1-4.

Trang Ho, A. S., 2006. *Tatoeba : Open collaborative multilingual "sentence dictionary",* Available at: <https://tatoeba.org/eng/> [Accessed 24 April 2021].

Treisman, A. M., 1964. Monitoring and storage of irrelevant messages and selective attention. *Journal of Verbal Learning and Verbal Behavior,* 3(6), pp. 449-459.

Tseng, J. H. W., 1987. Teaching well: A guide for undergraduate teaching-particularly for new engineering and technology faculty. *IEEE Transactions on Education,* 1(1), pp. 17-26.

Wu, C. T., Weissman, D. H., Roberts K.C., and Woldorff, M. G., 2007. The neural circuitry underlying the executive control of auditory spatial attention. *Brain Research,* 1134(1), pp. 187-198.