

Effects of acoustic and semantic cues on listening effort during native and non-native speech perception

Giulia Borghini^{a)} and Valerie Hazan

Department of Speech Hearing and Phonetic Sciences, Faculty of Brain Sciences, University College London, WC1N1PF London, United Kingdom

ABSTRACT:

Relative to native listeners, non-native listeners who are immersed in a second language environment experience increased listening effort and a reduced ability to successfully perform an additional task while listening. Previous research demonstrated that listeners can exploit a variety of intelligibility-enhancing cues to cope with adverse listening conditions. However, little is known about the implications of those speech perception strategies for listening effort. The current research aims to investigate by means of pupillometry how listening effort is modulated in native and non-native listeners by the availability of semantic context and acoustic enhancements during the comprehension of spoken sentences. For this purpose, semantic plausibility and speaking style were manipulated both separately and in combination during a speech perception task in noise. The signal to noise ratio was individually adjusted for each participant in order to target 50% intelligibility level. Behavioural results indicated that native and non-native listeners were equally able to fruitfully exploit both semantic and acoustic cues to aid their comprehension. Pupil data indicated that listening effort was reduced for both groups of listeners when acoustic enhancements were available, while the presence of a plausible semantic context did not lead to a reduction in listening effort.

© 2020 Acoustical Society of America. <https://doi.org/10.1121/10.0001126>

(Received 30 September 2019; revised 1 April 2020; accepted 2 April 2020; published online 8 June 2020)

[Editor: Tessa C. Bent]

Pages: 3783–3794

I. INTRODUCTION

Understanding a second language is a vital skill for those who live their life in a country where the main language is different from their native language. Because everyday interactions often occur in less than ideal listening conditions, non-native listeners may have to cope with the double challenge of an imperfect signal (e.g., during a telephone call) and an imperfect knowledge of the language (for a review see [Lecumberri et al., 2010](#)). It is precisely under suboptimal listening conditions that differences between native and proficient non-native listeners tend to emerge. Non-native listeners who show native-like performance in understanding spoken language in a quiet environment have been shown to be significantly less accurate than native listeners in the presence of noise or reverberation ([Mayo et al., 1997](#); [Takata and Nábělek, 1990](#)).

Behavioural studies evaluating the performance of native and non-native listeners indicate that individuals tend to adopt different strategies when attending to their first language (L1) in challenging listening conditions, as compared to their second language (L2). For example, relative to non-native listeners, native listeners generally rely more on lexical plausibility than on acoustic cues both when attending to intact speech and to speech in the presence of a competing talker ([Mattys et al., 2010](#)). Moreover, under heightened

cognitive load conditions (e.g., when simultaneously performing a visual search task), native listeners seem to further increase the weight given to contextual plausibility, while non-native listeners do not appear to show the same trend. Additionally, [Bradlow and Alexander \(2007\)](#) have shown that non-native listeners are able to benefit from modifications in speaking style, even when these are not combined with semantic cues. However, the same is not true for semantic cues; indeed, it was suggested that non-native listeners require a higher signal clarity in order to successfully exploit contextual cues.

Studies involving electroencephalography (EEG) further suggest that semantic integration is less straightforward during L2 compared to L1 speech processing. The N400 response is a negative event-related potential that has been linked to the ease of lexical access and integration, with greater N400 amplitude shown for more difficult words, or for words that are incongruent given the previous sentence context ([Lau et al., 2008](#)). [Hahne \(2001\)](#) found a delayed and more pronounced N400 response for coherent sentences in L2 compared to L1 listeners, suggesting that the semantic integration of the final word in the sentence was more difficult for listeners attending to a non-native language. However, despite being smaller in the non-native group, the difference between correct and incorrect sentences was significant in both native and non-native listeners. Similarly, [Song and Iverson \(2018\)](#) reported native listeners having a significantly greater N400 context-related difference (the difference between high and low predictability sentences) compared to non-native listeners. Taken together, the

^{a)}Also at: Cognition, Motion and Neuroscience Unit, Center for Human Technologies, Istituto Italiano di Tecnologia, Via E. Melen 83, 16152 Genova, Italy. Electronic mail: giulia.borghini.13@ucl.ac.uk, ORCID: 0000-0002-5232-0378.

evidence seems to indicate that lexical information is less readily exploitable for non-native compared to native listeners, and that its availability also depends on the listening environment that the individual is immersed in.

The use of a clear speaking style is another strategy aimed at enhancing intelligibility by making acoustic cues more readily available and more resistant to noise. Clear speech is an umbrella term used to indicate a speaking style that is spontaneously adopted by talkers when they are aware they may have difficulty being understood (Ferguson, 2004; Payton *et al.*, 1994). It is generally characterised by a decrease in speaking rate, a wider dynamic pitch range, an increase in intensity, an enlarged vowel space, an increase in the frequency and duration of pauses (Smiljanić and Bradlow, 2009). The size of the clear speech benefit varies significantly across different populations of talkers and listeners, also showing a high degree of flexible adaptation to listeners' needs (Hazan *et al.*, 2012). In a study from Bradlow and Bent (2002), non-native listeners showed only a small clear speech benefit: they improved significantly less than native listeners when presented with naturally produced clear speech as opposed to plain speech. This effect is due to the fact that while clear speech production involves some modifications that are likely to benefit all listeners, such as a lower speaking rate and wider pitch range, many of the adaptations made in clear speech are essentially "native listener oriented." Therefore, being able to take full advantage of them might require a native-like knowledge and experience of the phonology and phonetics of the target language. Indeed, in a further study, fluent and more language-proficient non-native listeners showed a large clear speech benefit when attending to sentences in noise (Smiljanić and Bradlow, 2011); they were also able to implement speech modifications that made them more easily intelligible to both native and non-native listeners.

Overall, experimental evidence reveals that listeners use a variety of intelligibility-enhancing cues to cope with challenging listening situations, and also that the advantages gained from such cues interact with one another and with the type of environment in which listeners are immersed. However, little is known about the implications of these speech perception strategies for listening effort. How does the presence of acoustic and contextual cues modulate cognitive effort during speech perception? Does the availability of semantic cues reduce listening effort by allowing easier access to sentence meaning, or conversely is the "contextual strategy" more effortful because of the need for higher level processing? Do those cues affect listening effort differentially in native compared to non-native listeners?

Recently, pupillometry has been deemed as a valid tool to study language processing load and listening effort, and measures of speech perception performance and effort have been included in studies as complementary to each other [see Zekveld *et al.* (2018) for an extensive review]. Pupillometry has been adopted as a technique to investigate changes in cognitive effort during speech perception attributable both to manipulations performed on the stimuli

presented and to different listeners' characteristics. In general, pupil response is shown to consistently increase when the intelligibility level of the speech presented decreases, unless extremely low levels of intelligibility are considered (Zekveld and Kramer, 2014; Zekveld *et al.*, 2010). Additionally, larger pupil responses have been reported when attending to semantically difficult words, and for low-frequency relative to high-frequency words (Chapman and Hallowell, 2015; Elshtain and Schaefer, 1968). Lexical manipulations such as increasing word frequency, reducing lexical competition, facilitating semantic processing and presenting stimuli with a sparser neighbourhood density have led to a reduction in pupil dilation (Kuchinsky *et al.*, 2013; Kuipers and Thierry, 2011; Schmidtke, 2014). Converging evidence also indicates that pupil response is sensitive to various types of manipulation at sentence level. Pupil diameter was found to increase when the syntactic complexity of sentences presented was higher, showing that syntactic complexity results both in greater and longer cognitive processing (Just and Carpenter, 1993; Schluhoff, 1982). Later studies showed similar findings, supporting the view that sentence complexity affects pupil response (Kramer *et al.*, 2012; Piquado *et al.*, 2010; Wendt *et al.*, 2016). A further study also indicated that attending to predictable sentences led to a more rapid reduction in listening effort compared to unpredictable sentences (Winn, 2016). Overall, pupil response has been shown to be a reliable index of listening effort during speech perception. Pupil dilation is specifically sensitive to sentence complexity, and emerging evidence also suggests that it may also reflect effort due to semantic integration.

Despite an increased interest in measuring listening effort, few studies have used pupillometry to investigate non-native speech comprehension. Overall, speech perception has been shown to elicit a greater pupil response when occurring in a non-native compared to a native language, both at the word level (Schmidtke, 2014) and for sentences presented in noise where intelligibility is equated across groups via a manipulation of the signal-to-noise-ratio (Borghini and Hazan, 2018). Importantly, higher English proficiency within bilingual listeners was associated with an earlier pupil response, and with a smaller effect of word frequency and neighbourhood density (Schmidtke, 2014).

The current study was designed to extend these findings by considering the effect of semantic context and acoustic cue enhancement on listening effort during the comprehension of spoken sentences. To this end, semantic plausibility and speaking style were manipulated both separately and in combination during a speech perception task in noise, which was administered to native and non-native listeners. As a result, either semantic enhancement, acoustic enhancement, or both were present. An adaptive procedure was used to match intelligibility levels across participant groups and conditions. Further, this study aimed at exploring how L2 proficiency level modulated listening effort in relation to the availability of semantic and acoustic cues. In summary, the present study was designed to answer the following research questions.

- (1) How does the presence of semantic cues affect listening effort in native and non-native listeners, at equated levels of intelligibility?
- (2) How does the presence of enhanced acoustic cues by means of a clear speaking style modulate listening effort in native and non-native listeners, at equated levels of intelligibility?
- (3) How does the combined presence of both clear speaking style and semantic cues impact on listening effort in native and non-native listeners?
- (4) How does language proficiency level affect listening effort in non-native listeners?

With respect to these research questions, two predictions were formulated. First, the availability of semantic cues would lead to an overall reduction in listening effort. This is based on previous findings showing that facilitations in semantic processing provided both at word and sentence level are associated with reduced pupillary response in native listeners (Ben-Nun, 1986; Chapman and Hallowell, 2015; Kuchinsky *et al.*, 2013; Kuipers and Thierry, 2011; Winn, 2016). Second, the reduction in pupil response due to semantic context would be smaller for non-native listeners, since they have been previously shown to be less able to exploit contextual cues at a sentence level (Bradlow and Alexander, 2007; Mattys *et al.*, 2010). In addition, we expected pupil response to be larger for non-native compared to native listeners, as found in Borghini and Hazan (2018). No prediction was made regarding the effect of clear speaking style on listening effort as there was no literature available on which to base such a prediction.

II. METHOD

A. Participants

Fifty five adults from two different language backgrounds participated in the study. One non-native participant was subsequently excluded due to performance for speech perception in noise exceeding two standard deviations of the mean performance in two out of four experimental conditions. The final non-native group included 35 participants (8 male and 27 female) with Italian as L1 and English as L2, aged 19–36 years ($M = 26.3$, $SD = 4.6$). All participants had been living in the United Kingdom for at least 3 months at the time of testing. The native group consisted of 19 native speakers of British English (10 male and 9 female), aged 19–34 years ($M = 25.8$, $SD = 4.8$). None of the participants reported suffering from cognitive or neural disorders, cataracts or diabetes, and they had not taken drugs or medications in the 48 h prior the experiment. These criteria were considered in order to exclude the potential effect of any confounding variable on pupil dilation. Participants were able to maintain the visual focus on the cross presented on the screen without contact lenses; glasses were allowed for those participants who needed them, and their usage did not interfere with the eye-tracker recording. The study was advertised via the University College London Psychology

subject pool and via social media; all participants provided written informed consent before participating in the study. They received monetary compensation for their time under a protocol approved by the Ethics Committee at UCL.

B. Stimuli and tests

1. Background tests

All participants had pure tone thresholds of 20 dB hearing level or better in both ears at octave frequencies between 250 and 8000 Hz. Non-native participants completed an online linguistic background questionnaire which collected information about their length of residence in the UK ($M = 2.2$ years, $SD = 2.0$, range: 0.25–7.5 years) and their overall self-reported English usage in percentage ($M = 45\%$, $SD = 17.7$, range: 10%–100%). The questionnaire was designed by adapting two existing set of questions: the Language History Questionnaire (Li *et al.*, 2014) and the Language Experience and Proficiency Questionnaire (Marian *et al.*, 2007). In addition, non-native participants were presented with the Listening module of the International English Language Testing System (IELTS), an international standardised test of English proficiency for non-native speakers (IELTS, 2017). This module comprised four sections, with ten questions in each section and a final score ranging from 0 to 40. The mean score obtained by participants was 26.7 ($SD = 8.8$, range: 5–38). The aim of this test was to obtain an accurate picture of the proficiency level of the participants, with a focus on their speech understanding skills.

2. Experimental stimuli

Semantic plausibility and the speaking style of the stimuli were manipulated for this study, resulting in the construction of four different sets of stimuli: semantically plausible and anomalous sentences produced in either a plain or clear speaking style. The Basic English Lexicon (BEL) sentence materials (Calandruccio and Smiljanic, 2012), including 20 lists of 25 sentences, were used as plausible stimuli, because the semantic context of each sentence was always consistent within the sentence itself. BEL sentences were specifically developed to test speech recognition for various listener populations, and contain lexical items and syntactic structures appropriate for use with non-native listeners. Examples of the sentences are: “The PARTY GAME was REALLY EASY,” “The OPERA THEATER is FULL this EVENING” (keywords in capital letters). Anomalous sentences were constructed using the following procedure. The keywords contained within each list of BEL sentences were categorised as nouns, verbs, adjectives, and adverbs, and subsequently shuffled across sentences within the same category and list. These anomalous sentences were then each manually checked by the experimenters and one additional native English speaker in order to ensure their grammaticality. Twenty lists of 25 anomalous sentences retaining the same grammatical structure of the BEL were thus created and used as anomalous stimuli. Examples of

these sentences are “The VEGETABLES OPEN a DIFFICULT HAT,” “The UGLY BOOKS SAW the FATHER” (keywords are in capital letters). Each plausible and anomalous sentence had four keywords, which were used to score comprehension.

Four native Southern British English speakers (two females) were recorded producing these sentences in a sound-attenuated booth in the Department of Speech, Hearing and Phonetic Sciences at UCL, with a sampling rate of 44.1 kHz, 16-bit resolution. Four speakers were recorded in order to increase the ecological validity of the study and improve the generalisability of findings, by ensuring that listeners were not unduly influenced by the individual characteristics of a single speaker. Speakers read the sentences from a monitor and were asked to produce each set of sentences (plausible and anomalous) twice. For the plain speaking style, they recorded the sentences in a conversational self-paced rate, without any particular focus on clarity. To elicit stimuli in a clear speaking style, talkers were instructed to pronounce sentences as if they were speaking to a person with hearing impairment, or with a second language learner, similarly to the procedure used by Bradlow and Alexander (2007). Recordings were root-mean-square (RMS) normalised to an average amplitude of 65 dB. Average sentence duration was 1.9 s (SD = 0.2) for plain stimuli and 4.7 s (SD = 1.6) for clear stimuli.

3. Experimental task

The experimental task consisted of a speech intelligibility test: participants were asked to listen to sentences and to repeat them back to the experimenter. Sentences were presented via a loudspeaker placed in front of the participant at a distance of approximately 1 m, in order to ensure participants’ comfort and avoid pupil measurement being affected by potential discomfort caused by headphones. The experimental task consisted of two sections: one performed in quiet and the other with background noise. Each of the two sections included four parts where semantic plausibility and speech clarity were manipulated, resulting in four conditions: plausible and semantically anomalous sentences pronounced both in plain and clear speaking style. The four experimental conditions were labelled as *plaus_plain* (PP), *plaus_clear* (PC), *anom_plain* (AP), and *anom_clear* (AC). The test in quiet was always presented first, while, within each of the two sections, the presentation order of the four conditions was randomised. Pupil data from the test in quiet were not considered for analyses. The sound level for the entire duration of the experiment was constant at ~67–69 dB, as measured by a sound level meter, to avoid any confounding effect on the pupil dilation due to changes in the sound intensity. Each participant was presented with sentences pronounced by a single speaker across all experimental conditions, randomly selected among the four speakers recorded. This was done to reduce performance variability within the same participant; in this way, the adaptive procedure adopted for determining the signal-to-noise

level to be used for the speech intelligibility in noise conditions could function optimally. Over the whole experimental session, each participant was presented with 220 sentences, with each sentence only played once for a given participant. The same list was never repeated across the semantically plausible and anomalous conditions for each participant.

a. Speech perception in quiet. Participants were presented with six practice items, followed by one block of 10 sentences each per experimental condition (*plaus_plain*, *plaus_clear*, *anom_plain*, and *anom_clear*) presented in quiet. The purpose of this test condition is twofold. First, it is useful to get an evaluation of the participants’ performance level in quiet. Additionally, these data were used to check whether the use of stimuli recorded by different speakers had an effect on speech intelligibility.

b. Speech perception in background noise. For each of the four conditions, two experimental blocks were presented. In the first block, which included 25 sentences, an adaptive procedure (Levitt, 1971) was used to estimate the signal-to-noise ratio (SNR) required to reach the target intelligibility level of 50% correctly reported keywords. 50% was chosen as target intelligibility level in order to elicit a significant pupil dilation without resulting in participants giving up the task due to extremely poor performance (Zekveld and Kramer, 2014). An eight-talker babble noise obtained from recordings of spontaneous speech from four female and four male English native speakers was used as a masker for the speech. The SNR was manipulated by adapting both the target and the masker levels, so that the overall intensity of the compound signal remained unchanged at 67–69 dB. The first sentence was always presented at 20 dB SNR; in the following sentences, the SNR was manipulated to target the level at which half of keywords were correctly reported. The changes in step size were defined by an algorithm according to the participant’s performance and test stage; steps of 9 dB were applied during the initial stage and smaller 3 dB steps closer to the target SNR. Following this adaptive procedure, the SNR values corresponding to the reversals were averaged to obtain a single SNR value. In the following block, 20 sentences were presented using the fixed SNR level previously determined for this participant. The same design consisting of one adaptive (25 trials) + one fixed (20 trials) block was repeated for each of the four experimental conditions.

C. Procedure

Prior to the test session, non-native participants completed the language background questionnaire; the audiometric test was then performed at the beginning of the session. During the test, participants sat in a comfortable chair in a sound-attenuated booth, in front of a screen positioned 70 cm away. A headrest supporting the forehead and chin of the participant was used in order to reduce movement artefacts while performing the experiment. The

luminance of the room ranged from 65 to 110lx, and was individually adjusted to enable the pupil of the participant to be approximately in the middle of its dynamic range, in order to prevent ceiling and floor effects. A nine-point calibration procedure was initiated and validated. Then, the experimental task started and participants were instructed to maintain their gaze and focus at a cross at the centre of the screen, in order to facilitate reliable pupil size measurements. The screen had a grey background colour that did not change for the entire duration of the experiment to avoid abrupt changes in luminance level. Each trial began with the fixation cross turning black, and with 2 s of silence or babble noise (in the test in quiet and in noise, respectively) prior to sentence onset, in order to determine the baseline pupil diameter. The sentence was then played, and the fixation cross remained black for three additional seconds following the sentence offset, to allow enough time for the pupil to reach its maximum dilation before the repetition prompt appeared. The fixation cross turning green was the signal for the participant to repeat the sentence back to the experimenter who was simultaneously scoring keyword accuracy on another screen. Participants were told that they could rest their eyes and shift their gaze while the fixation cross was green between one trial and the other. The experimenter then initiated the following trial after making sure that the participant was ready to continue. A break was taken preferably at the end of each block, but pauses at any time between trials were also allowed in case participants needed to rest their eyes. At the end of the speech perception test, non-native participants were presented with the Listening section of the IELTS. The audio tracks were presented from the loudspeaker, and participants were asked to write their answers on printed forms.

D. Pupillometry

During the speech perception test, the pupil size and location of the left eye were measured using an EyeLink 1000 eye-tracker. The system uses infrared video-based tracking technology, with a spatial resolution of ~ 0.01 mm (value calculated for a pupil diameter of 5 mm), and was positioned at a horizontal distance of 70 cm from the participant. Pupil data were collected at the sampling rate of 500 Hz, and stored on a linked PC. At the end of each experimental session, an artificial pupil of known diameter was recorded for 3 s. The original pupil signal of the EyeLink 1000 was expressed in arbitrary units (SRResearch, 2009); the pupil signal was then converted to millimetres by an off-line calibration. During data collection, the experimenter was able to visually inspect the video recording from a monitor and to take action if needed (e.g., reminding the participants to fixate the centre of the screen, asking them to move in order to have the pupil in the eye-tracker searching area). The experimental task and data collection were controlled using a script in MATLAB version R2015a.

The pupil data were pre-processed following a multi-step procedure. Sequences of missing values corresponding

to blinks were expanded asymmetrically such that the time window including 80 ms prior to and 160 ms following a missing data stretch was omitted, in order to exclude local pupil size disturbances caused by partial eyelid closures (Winn *et al.*, 2018). Linear interpolation was performed to replace missing data corresponding to blinks. A smoothing first-order 10 Hz low-pass filter was applied in order to reduce the high frequency noise in the pupil data. The one second period preceding the stimulus onset was regarded as baseline. A minimum of 25 samples (corresponding to 50 ms) was defined as a requirement for a valid baseline estimation: otherwise, the baseline window was extended backward until the criterion was met.

Following the pre-processing, pupil data were averaged separately for each participant and condition (plaus_plain, plaus_clear, anom_plain, and anom_clear). The change in pupil diameter was quantified in mm relative to the baseline. The time window considered for the analyses started 1 s prior the sentence onset, and ended 4.9 and 7.7 s after the sentence offset for plain and clear stimuli, respectively. This choice was made to consider solely the time window before participants were prompted to repeat back the sentence they heard (the prompt always appeared 3 s after the sentence offset). The rationale for it was to exclude changes in the pupil diameter caused by movement planning and execution (Richer and Beatty, 1985). Two pupil outcome measures were obtained from the average trace of each participant and condition: Mean pupil dilation relative to baseline pupil diameter (time windows considered: between 0 and 4.9 s after the stimuli onset for plain stimuli, between 0 and 7.7 s after the stimuli onset for clear stimuli) and Peak pupil dilation, (maximum positive deviation from the baseline during the 4.9 and 7.7 s following stimuli presentation respectively for plain and clear stimuli).

E. Statistical analyses

One way repeated-measures analyses of variance (ANOVAs), one-way multivariate analysis of variance (MANOVA), and mixed design ANOVAs were conducted to test whether test condition (semantic plausibility and speaking style) and linguistic background of participants (native or non-native listener) affected behavioural data (intelligibility and SNR levels). Stepwise regressions were performed to test whether data collected about L2 proficiency and usage were correlated with listeners' ability to tolerate background noise.

Mixed-effect regression models were performed to analyse the effect of semantic plausibility, speaking style, and language background on pupil measures. Additionally, mixed-effect regression models were also used to investigate the effect of length of residence, overall English use and IELTS score on pupil measures for non-native listeners only.

III. RESULTS

A. Behavioural results

Intelligibility scores in quiet for each of the four conditions are displayed in Fig. 1, left panel.

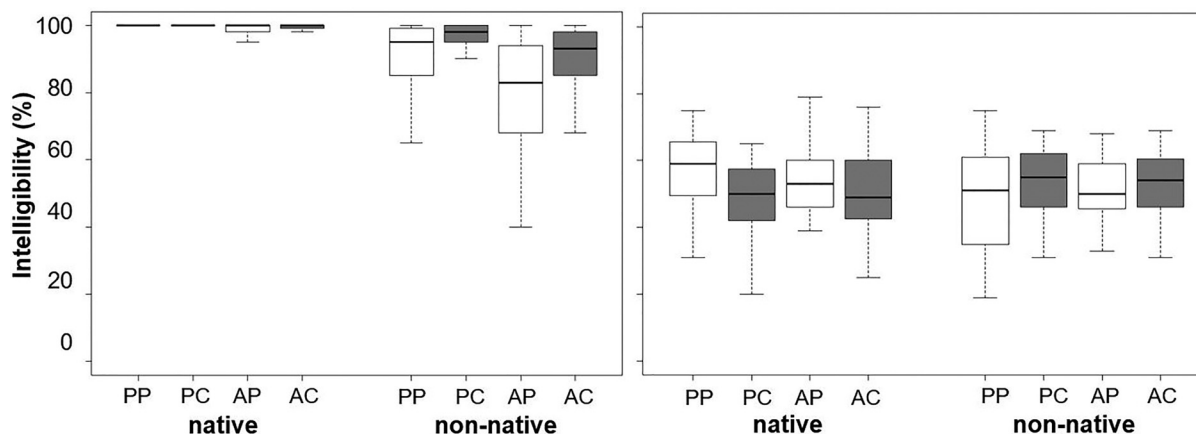


FIG. 1. Keyword intelligibility rates for sentences presented in quiet (left panel) and in noise at SNR levels individually set to target a 50% intelligibility level (right panel). Intelligibility rates are shown as a function of test condition and participant group (PP = plausible-plain, PC = plausible-clear, AP = anomalous-plain, AC = anomalous-clear). The ends of the whiskers represent one and a half times the interquartile range.

A one-way MANOVA was run to determine whether there was an overall difference in performance between language groups in quiet across the four conditions. Pillai's criterion was used to assess the statistical significance, because it is more robust for unequal sample sizes than the standard Wilks' λ (Tabachnick *et al.*, 2007). There was a statistically significant difference between native and non-native listeners on the combined dependent variables [$F(4, 51) = 6.00, p < 0.001$; Pillai's Trace = 0.616, partial $\eta^2 = 0.32$]. Overall, native listeners had better performance in quiet compared to non-native listeners.

As the performance of native participants was at ceiling (>98%), further analyses focused on data for non-native participants. A 2×2 repeated measures ANOVA evaluated the effect of semantic plausibility and speaking style on non-native listeners' performance in quiet. There was a statistically significant interaction between semantic plausibility and speaking style [$F(1, 34) = 9.93; p = 0.003$]. Main effects of semantic plausibility and speaking style were also found to be significant [$F(1, 34) = 52.71; p < 0.001$ and $F(1, 34) = 40.41; p < 0.001$, respectively]. Follow-up Bonferroni-adjusted pairwise comparisons indicated that both in plausible and anomalous sentences, clear speech led to higher intelligibility rates compared to plain speech ($p < 0.001$). Additionally, in both speaking style conditions, intelligibility rates were higher for plausible sentences than for anomalous sentences ($p < 0.001$). Last, for each condition (plausible and anomalous sentences pronounced in a plain and clear speaking style) an independent one-way ANOVA was performed on non-native listeners' performance in order to check that the use of four different speakers in the experiment did not significantly affect intelligibility levels. In all conditions in quiet the effect of the speaker on the percentage of keywords correctly reported by non-native listeners did not reach significance level, hence individual characteristics of the speaker did not affect intelligibility levels in quiet.

Figure 1, right panel, displays intelligibility rates during the speech perception test in noise for each experimental condition. The adaptive block used to set SNR level is not

included in the analyses. A mixed design ANOVA performed on intelligibility levels with semantic plausibility (plausible and anomalous) and speaking style (plain and clear) as within-subjects factors, and language background (native and non-native) as a between-subjects factor was performed in order to verify that the adaptive procedure was successful in equating intelligibility levels across conditions and groups at 50%. Results showed a significant interaction between speaking style and participants' language group [$F(1, 52) = 6.12, p = 0.017$]. Follow-up Bonferroni-adjusted pairwise comparisons indicated that marginally higher intelligibility scores were obtained with clear speech as compared to plain for native listeners ($p = 0.050$), and that marginally higher intelligibility scores were obtained for native than for non-native participants when attending to clear speech ($p = 0.052$). Despite the implementation of an adaptive procedure, behavioural results were not perfectly equated. Importantly, however, only a small degree of variation is left across language groups and conditions.

Figure 2 displays the corresponding SNR levels at which the fixed procedure blocks in noise were run. As expected, the SNR levels corresponding to 50% intelligibility were significantly lower (i.e., showed greater tolerance of noise) for native participants than for non-native participants [$F(1, 52) = 29.58, p < 0.001$], for plausible compared to anomalous stimuli [$F(1, 52) = 63.52, p < 0.001$] and for sentences produced in a clear compared to a plain speaking style [$F(1, 52) = 337.04, p < 0.001$]. None of the interactions were found to be significant.

1. Proficiency effect

Individual differences in SNR levels were investigated in order to test whether second language proficiency and language usage (as evaluated from the language background questionnaire) were correlated with listeners' ability to tolerate background noise. Stepwise regression analyses were performed for each of the four experimental conditions (plaus_plain, plaus_clear, anom_plain, and anom_clear).

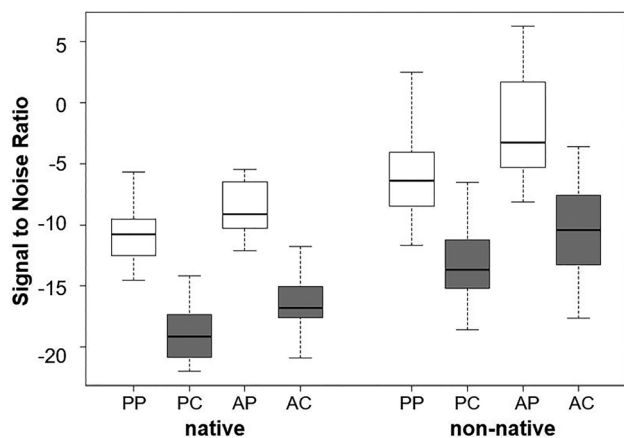


FIG. 2. SNR values targeting 50% intelligibility for each experimental condition for native and non-native listeners (PP = plausible-plain, PC = plausible-clear, AP = anomalous-plain, AC = anomalous-clear). The ends of the whiskers represent one and a half times the interquartile range.

Length of residence, overall self-reported English use and score for the Listening section of the IELTS were entered as predictors. The variance of inflation factor was smaller than 2 for each regression coefficient considered; therefore, we can assume that the regression results were not affected by multicollinearity. Results showed that the score obtained for the IELTS Listening test significantly contributed to predicting the estimated SNR level for all the experimental conditions: a higher score predicted a lower SNR (i.e., better performance). Table I summarises the results for all the significant predictors reported above.

B. Pupil results

Descriptive statistics for the pupil data collected during the sentence blocks in noise with individually fixed SNR are reported in Table II. Measures include mean and peak pupil dilation over the baseline following stimulus presentation.

1. Effect of semantic plausibility, speaking style and language background on listening effort

Pupil dilation curves for native and non-native listeners in all experimental conditions are displayed in Fig. 3. The effect of semantic plausibility, speaking style and language background on pupil measures was tested using a model comparison approach. A series of mixed-effects regression models was implemented using maximum likelihood estimation with the lme4 package in the R environment (Bates et al., 2014; R Core Team, 2017).

Separate models were built for mean and peak pupil dilation. The initial saturated model included interaction terms for all independent variables as fixed effects with by-subject random intercepts and slopes (Barr et al., 2013). Due to non-convergence, the random effect structure was simplified to include by-subject random intercepts but no random slopes. The resulting converged maximal models for both variables (mean and peak pupil dilation) included the following fixed effects: language background (2: native and non-native), semantic context (2: plausible and anomalous), and speaking style (2: plain and clear). The maximal model included up to three-way interactions between language background, semantic plausibility, and speaking style. A backward selection procedure (moving from higher-level interaction terms to single term) was adopted to remove fixed terms that did not significantly improve model fit. Model residuals via chi-square tests ($\alpha = 0.05$) were compared from the most complex models (containing the largest interaction term) to the least complex models (containing only single terms as fixed effects and participant as random effect). If an interaction term was significant, all lower level effects involved in the interaction were included in the final model.

The final models for both mean and peak pupil dilation included fixed effects of language background and speaking style (but no interaction), as well as by-subject random intercepts. P-values for each term included in the final models were estimated from z-score using normal distribution (see Tables III and IV for coefficients and p values for the mean and peak pupil dilation, respectively). These results indicate that overall both mean and peak pupil measures were greater for non-native compared to native listeners, and for sentences produced in plain compared to clear speaking style. However, there was no significant effect of semantic plausibility on mean and peak pupil dilation.

2. Non-native listeners

In order to investigate the effect of length of residence, overall English use and IELTS score on pupil measures, the same type of analysis was run with non-native listeners only for mean and peak pupil dilation. The resulting converged maximal models for both variables included the following fixed effects: semantic plausibility (2: plausible and anomalous), speaking style (2: plain and clear), length of residence, self-reported English use, and IELTS score. The random effects structure included by-subject random intercepts but no random slopes. The maximal model only included up to two-way interactions between conditions

TABLE I. Effect of individual differences on the SNR levels for non-native listeners, stepwise regression results.

Dependent variable	Predictor	R ²	B	Std. error	Std. beta	F	t	Sig.
SNR plaus_plain	IELTS Listening	0.695	-0.485	0.056	-0.834	75.239	-8.674	<0.001
SNR plaus_clear	IELTS Listening	0.514	-0.349	0.059	-0.717	34.842	-5.903	<0.001
SNR anom_plain	IELTS Listening	0.593	-0.691	0.100	-0.770	48.025	-6.930	<0.001
SNR anom_clear	IELTS Listening	0.545	-0.422	0.067	-0.738	39.558	-6.290	<0.001

TABLE II. Descriptive statistics of the pupil measures in noise.

Pupil data in noise						
Plausible stimuli						
Speaking style:	Plain			Clear		
Pupil outcome	All M (SD)	Non-native M (SD)	Native M (SD)	All M (SD)	Non-native M (SD)	Native M (SD)
Mean dilation, mm	0.12 (0.17)	0.15 (0.18)	0.07 (0.14)	0.05 (0.17)	0.10 (0.16)	-0.05 (0.16)
Peak dilation, mm	0.28 (0.24)	0.33 (0.24)	0.20 (0.22)	0.19 (0.20)	0.25 (0.20)	0.08 (0.14)
Anomalous stimuli						
Speaking style:	Plain			Clear		
Pupil outcome	All M (SD)	Non-native M (SD)	Native M (SD)	All M (SD)	Non-native M (SD)	Native M (SD)
Mean dilation, mm	0.09 (0.16)	0.14 (0.14)	0.00 (0.16)	0.05 (0.21)	0.09 (0.18)	-0.03 (0.23)
Peak dilation, mm	0.25 (0.22)	0.31 (0.21)	0.13 (0.19)	0.21 (0.23)	0.26 (0.23)	0.13 (0.20)

(semantic plausibility and speaking style), and the background measures considered (length of residence, self-reported English use, and IELTS score). An identical procedure as described above was applied for model comparison.

The final models for both pupil dilation measures only included the fixed effect of speaking style, confirming an overall reduction of mean and peak pupil dilation for non-native listeners for sentences produced with clear compared

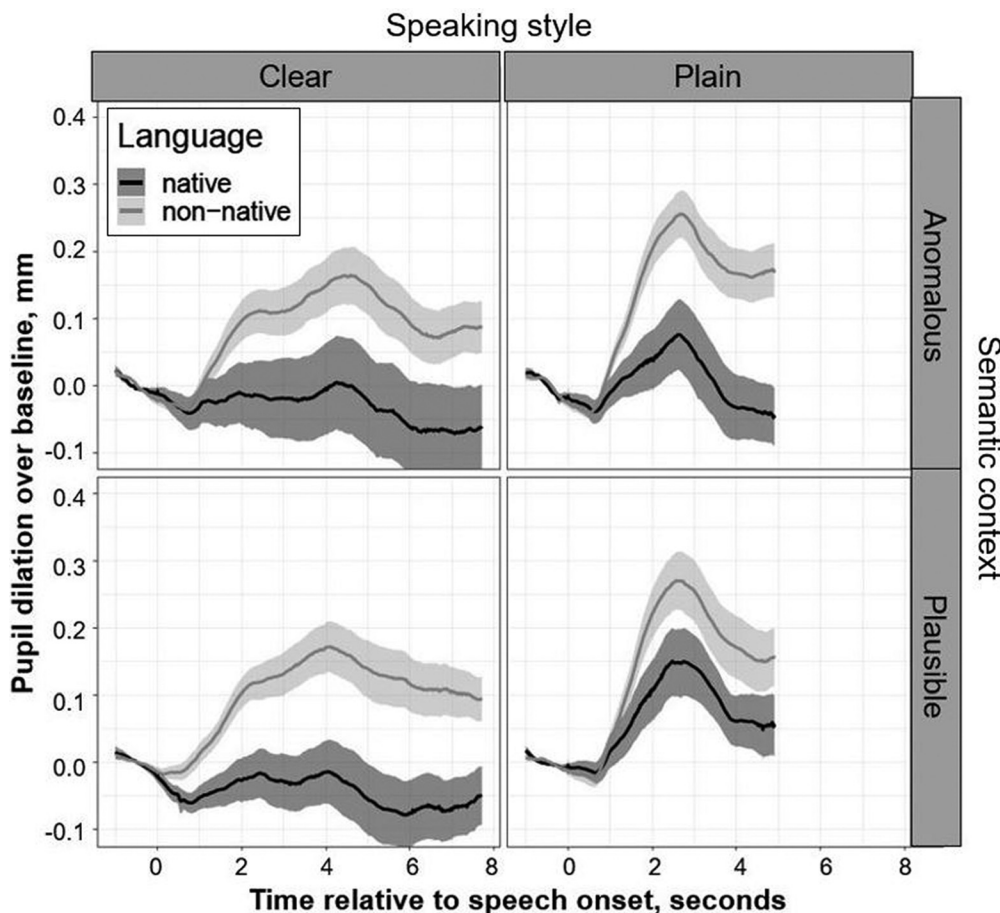


FIG. 3. Mean pupil response over time during speech perception in all experimental conditions, for native and non-native listeners. The coloured ribbon represents standard error.

TABLE III. Fixed and random effects in a mixed-effects model of mean pupil dilation for all listeners. Note: Number of observations = 216; participants (N) = 54.

Mean ~ Language + Style + (1 Participant)					
Fixed effects		Estimate	SE	t	Estimated p
(intercept)		-0.03	0.03	-0.85	0.40
Language	Non-native	0.12	0.04	2.85	0.004
Style	Plain	0.06	0.01	4.38	< 0.001
Random effects		Variance			
Participant	(intercept)			0.02	
Residual				0.01	

to plain speaking style (see Tables V and VI for coefficients and estimated *p* values for mean and peak pupil dilation, respectively). Overall, analyses on the non-native pupil data did not reveal any significant effect of length of residence, self-reported English use and IELTS score on pupil measures.

IV. DISCUSSION

The current experiment was designed to investigate how the presence of semantic and acoustic enhancements affect listening effort in native and non-native listeners when levels of intelligibility are equated. Behavioural results showed that native listeners were able to tolerate a less favourable SNR compared to non-native listeners, when the intelligibility level across the two groups was equated. Additionally, both native and non-native listeners were able to cope with a lower SNR level when attending to semantically plausible compared to anomalous stimuli, and when sentences were produced with a clear compared to a plain speaking style. The two groups of listeners hence differed in their overall speech perception ability in noise. Importantly however, native and non-listeners were equally able to take advantage of acoustically enhanced speech (i.e., speech produced in a clear speaking style). In addition, the availability of contextual information was shown to be effective in improving the intelligibility of speech in noise for both listener groups (Drager and Reichle, 2001; Wingfield and Tun, 2007; Zekveld et al., 2011).

TABLE IV. Fixed and random effects in a mixed-effects model of peak pupil dilation for all listeners. Note: Number of observations = 216; participants (N) = 54.

Peak ~ Language + Style + (1 Participant)					
Fixed effects		Estimate	SE	t	Estimated p
(intercept)		0.10	0.04	2.46	0.014
Language	Non-native	0.15	0.05	3.00	0.003
Style	Plain	0.06	0.02	3.99	< 0.001
Random effects		Variance			
Participant	(intercept)			0.03	
Residual				0.01	

TABLE V. Fixed and random effects in a mixed-effects model of mean pupil dilation for non-native listeners. Note: Number of observations = 140; participants (N) = 35.

Mean ~ Style + (1 Participant)					
Fixed effects		Estimate	SE	t	Estimated p
(intercept)		0.09	0.02	3.74	< 0.001
Style	Plain	0.05	0.02	2.92	0.003
Random effects		Variance			
Participant	(intercept)			0.02	
Residual				0.01	

The results further establish that a clear speaking style is an effective means of enhancing speech perception for both native and non-native listeners (Bradlow and Bent, 2002; Gryn timer et al., 2011; Van Engen et al., 2014). Nevertheless this pattern of results is somewhat in contrast with previous research that reported a greater clear speech benefit for native compared to non-native listeners (Bradlow and Bent, 2002), and a contextual benefit for non-native listeners only when speech clarity was simultaneously enhanced (Bradlow and Alexander, 2007). However, the perspective taken here is slightly different from that in previous studies. Indeed, rather than measuring how acoustic and semantic cues improved intelligibility, in the current experiment intelligibility remained fixed at 50% and changes in the SNR levels were compared. However, because lowering the SNR means enhancing the level of energetic masking, it can be argued that performing the task with a less favourable SNR further reduced the availability of relevant acoustic cues, and that this further reduction would act in addition to the experimentally planned manipulation on signal clarity. Therefore, while the speaking style benefit showed by native and non-native listeners was about the same magnitude (approximately 8 dB), the native group might have experienced a greater degradation of the target speech due to the combined effect of a louder background noise and of the speaking style manipulation. It might thus remain true that native listeners are able to gain a greater advantage from clear speaking style compared to non-native listeners, as previously reported (Bradlow and Bent, 2002).

Results from pupil data confirmed that listening effort during sentence perception is higher for non-native compared to native listeners when intelligibility levels are

TABLE VI. Fixed and random effects in a mixed-effects model of peak pupil dilation for non-native listeners. Note: Number of observations = 140; participants (N) = 35.

Peak ~ Style + (1 Participant)					
Fixed effects		Estimate	SE	t	Estimated p
(intercept)		0.26	0.03	7.63	< 0.001
Style	plain	0.06	0.02	3.16	0.002
Random effects		Variance			
Participant	(intercept)			0.03	
Residual				0.01	

equated, and this effect was found regardless of which contextual and acoustical enhancements were made available to the listener. These findings are in line with previous research evaluating listening effort in non-native listeners during spoken word recognition (Schmidtke, 2014) and sentence perception (Borghini and Hazan, 2018).

Surprisingly, the presence of a coherent semantic context within a sentence was not found to reduce listening effort, as measured by a reduction in pupil dilation. This result was unexpected, given a study showing a consistent reduction in cognitive effort for predictable versus unpredictable sentences in native listeners (Winn, 2016). However, the speech material used to manipulate context plausibility in the present study was inherently different from the material used in Winn (2016). Indeed, the listening task used in Winn (2016) featured sentences with high and low semantic context (e.g., “*Stir your coffee with a spoon*” versus “*Jane thought about a spoon*”), while in the study presented here the plausible sentences were contrasted against semantically anomalous sentences (e.g., “*The talented artist drew a picture*” vs “*The vegetables open a difficult hat*”). Moreover, since sentences were presented in blocks of either semantically plausible or anomalous sentences, listeners could have anticipated if a semantically coherent or incoherent sentence was expected. Listeners may have been able to assess whether it would be advantageous to exploit semantic cues or not, given the lack of a coherent semantic context in the anomalous condition. By ruling out the use of a semantic strategy, they may have consequently inhibited the processing itself of the semantic information available in the sentence. This reasoning could explain why a theoretically more effortful meaning integration (as required for anomalous compared to plausible sentences) did not elicit a greater pupil response, as found by Kuipers and Thierry (2011) when considering unrelated versus related word/picture pairs. Therefore, the stimulus selection adopted in the present study might not have been ideal to the purpose of investigating whether non-native listeners are able to benefit from semantic context in terms of listening effort reduction. One additional factor that could have contributed to the lack of effect of semantic context on the pupil measurements is that the SNR for plausible stimuli was overall less favourable than the SNR for anomalous stimuli. This difference itself could have made listening more effortful and thus counteract any difference in listening effort due to semantic context. On the other hand, however, equating SNR across participants would have led to differences in the intelligibility levels, therefore causing a confounding effect in pupil response, since intelligibility itself has been shown to affect pupillometric measures of listening effort (Zekveld and Kramer, 2014; Zekveld et al., 2010).

This study also showed a reduced listening effort attributable to a clear speech benefit. There is limited availability of research addressing the effects of speaking style on pupil response. Simantiraki et al. (2018) similarly found reduced listening effort for sentences pronounced with a clear (author referred to this condition as “Lombard,” since

stimuli were recorded in the presence of background noise) compared to a plain speaking style when presented at -3 and -5 SNR (but not at -1 SNR). However, intelligibility levels across speaking styles were not equated, therefore a direct comparison with the current study might be misleading. Along a related line of investigation, Koch and Janse (2016) did not find any speech rate effect on pupil response. To enhance the ecological validity of their study, the authors did not perform any artificial time-compression of speech, but instead used conversational materials with a natural variation in speech rate. Although changes in the speaking rate are a frequent component of speaking style adaptations that fall under the term of “clear speech,” other factors that are intrinsic to a clear speaking style may have contributed to the significant reduction in listening effort for clear compared to plain speaking style reported in the present study. The combined contribution of different clear speech strategies may have affected the ease of processing and the cognitive demand in a way that solely a reduction in the speech rate did not. Additionally, it is worth noting that in the current research a reduction of listening effort linked to the use of clear speaking style was found despite an overall more challenging SNR level in the clear compared to the plain speech condition. Indeed, listeners were able to tolerate a less favourable SNR when attending to clear compared to plain speech at equated levels of intelligibility.

One additional focus of the present study was the investigation of the effects of L2 proficiency on non-native listening effort and performance. Non-native listeners who obtained higher IELTS Listening score were able to tolerate a greater degree of background noise for all the experimental conditions. Previous research showed that highly proficient non-native listeners were better at ignoring semantic L1 interference during a listening task (L1 words presented as distractors during the task), demonstrating that they were better able to selectively attend to the target language compared to less proficient listeners (Elston-Güttler and Gunter, 2009). One possibility is that an increased interference from L1 vocabulary contributed to the lower performance (i.e., increased SNR) for less proficient non-native listeners. Results from the present study are somewhat in contrast with findings from Francis et al. (2018), which reported a positive correlation between pupil dilation and English proficiency score in a study exploring speech perception in Dutch listeners attending to sentences in both Dutch and English. However, the authors found a significant correlation only in few of the experimental conditions examined (2 out of 8), also including trials where native speech perception was considered. Therefore, the authors suggested that the results might simply reflect the influence of a more general cognitive capacity, which also affects susceptibility to distraction. Additionally, as noted when discussing the effect of semantic plausibility, the lack of effect of background measures of proficiency on the pupil measurements might be partially attributable to the fact that non-native listeners with better IELTS scores attended to sentences presented at a less favourable SNR compared to less proficient

non-native listeners. However, the technique of pupillometry itself may not have been ideal to uncover individual differences. Indeed, while pupillometry has been deemed to be a reliable technique to uncover between subjects differences at a group level, pupillometric data from a single person are not considered as reliable due to a wide range of individual variability (Winn *et al.*, 2018). Visual inspection of the pupil data collected for this study also confirmed the large individual variability of the pupil measures. In order to clarify the relationship between listening effort and second language proficiency, further research might better tailor the distinction between high and low proficiency in non-native listeners. A recruitment process targeted at selecting two groups of participants with distinct proficiency levels may be a better solution for exploring the impact of proficiency on listening effort.

Last, results from the current study also contribute to enriching our general understanding of the mechanisms affecting cognitive load during speech perception. Various models in this regard have been proposed and discussed in the literature (Pichora-Fuller *et al.*, 2016; Rönnberg *et al.*, 2019; Rönnberg *et al.*, 2013; Rönnberg *et al.*, 2008). However, their primary focus has been on the cognitive mechanisms that come into play when listeners are attending to an impoverished signal (due to perceptual deficits or hearing aids), to speech masked by noise, or a combination of the two (Pichora-Fuller *et al.*, 2016). Very limited attention is given to the specific challenges encountered by listeners attending to a non-native language. Nevertheless, some challenges are certainly shared between hearing-impaired listeners and those who are attending to a non-native language. The most recent version of the Ease of Language Understanding (ELU) model (Rönnberg *et al.*, 2019) predicts that the degradation of the language signal increases the probability of a mismatch between the input stream of language and the stored mental representations of phonological and lexical units. According to the ELU model, an explicit and deliberate involvement of working memory (WM) to solve the mismatch is key to a successful language comprehension in challenging conditions, but it is also considered to be linked to an increase in cognitive effort. In the present study, speech clarity was directly manipulated. By its very nature, clear speech is meant to facilitate listeners in the process of phoneme recognition, and therefore to reduce the occurrence of mismatches between input signal and stored phoneme representations. By applying the same rationale as above, clear speech is supposed to reduce the explicit reliance of speech understanding on WM and cognitive functions, and therefore a reduced cognitive load would be predicted by the ELU model when listeners are attending to clear compared to plain speech. Results confirmed a reduced listening effort when native and non-native listeners are attending to clear speech, thus corroborating the predictions of the ELU model for both listener groups. Importantly, being able to expand the validity of the predictions of cognitive load based on the ELU model beyond the consideration of native normal hearing and hearing impaired listeners further demonstrates the theoretical reliability of the model.

However, based on the ELU model, a reduced listening effort when participants are attending to semantically plausible sentences would be predicted. Indeed, Rönnberg *et al.* (2019) reason that the availability of a coherent semantic context leads to a reduced reliance on cognitive “repair” functions such as WM, because the probability of mismatches is reduced by constraining the number of candidate words/phonemes. On the contrary, results from our study did not indicate a reduction in listening effort in the presence of an enhanced semantic context. A more tailored contrast between experimental conditions (a comparison between high and low availability of semantic context, as opposed to the use of plausible and anomalous stimuli) might help to clarify whether a reduction in listening effort can be achieved through the exploitation of semantic content. Finally, in its current structure, the ELU model is ineffective in accounting for the additional challenges faced by non-native listeners. A useful addition to expand the applicability of the ELU model would be the inclusion of a “language-specific knowledge” component feeding as a support into the pre-diction function, implying that native-like language knowledge allows for fast and implicit prediction making, but would also account for any increase in the processing load due to a reduced proficiency in L2. Moreover, the consideration of potential L1 interference would further improve the effectiveness of the model in generating realistic predictions for a wider range of listeners.

In conclusion, the study reported in this paper confirmed that pupil response is a sensitive measure to uncover listening effort differences between native and non-native listeners, and confirmed previous findings (Borghini and Hazan, 2018) that a greater listening effort is required when trying to understand a second compared to a native language in noise even when overall intelligibility is matched. In addition, this research uncovered potential effects of speaking style on listening effort: attending to sentences produced in a clear speaking style has a reduced cost in terms of cognitive effort both for native and non-native listeners.

ACKNOWLEDGMENTS

This work was supported by the Economic and Social Research Council (Grant No. ES/J500185/1). The authors would like to thank Outi Tuomainen for her statistical support.

- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). “Random effects structure for confirmatory hypothesis testing: Keep it maximal,” *J. Mem. Lang.* **68**(3), 255–278.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). “Fitting linear mixed-effects models using lme4,” [arXiv:1406.5823](https://arxiv.org/abs/1406.5823).
- Ben-Nun, Y. (1986). “The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing,” *Brain Lang.* **28**(1), 1–11.
- Borghini, G., and Hazan, V. (2018). “Listening effort during sentence processing is increased for non-native listeners: A pupillometry study,” *Front. Neurosci.* **12**, 152.
- Bradlow, A. R., and Alexander, J. A. (2007). “Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners,” *J. Acoust. Soc. Am.* **121**(4), 2339–2349.
- Bradlow, A. R., and Bent, T. (2002). “The clear speech effect for non-native listeners,” *J. Acoust. Soc. Am.* **112**(1), 272–284.

- Calandruccio, L., and Smiljanic, R. (2012). "New sentence recognition materials developed using a basic non-native English lexicon," *J. Speech Lang. Hear. Res.* **55**(5), 1342–1355.
- Chapman, L. R., and Hallowell, B. (2015). "A novel pupillometric method for indexing word difficulty in individuals with and without aphasia," *J. Speech Lang. Hear. Res.* **58**(5), 1508–1520.
- Drager, K. D., and Reichle, J. E. (2001). "Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: Applications for AAC," *J. Speech Lang. Hear. Res.* **44**(5), 1052–1057.
- Elshtain, E. L., and Schaefer, T. (1968). "Effects of storage load and word frequency on pupillary responses during short-term memory," *Psychon. Sci.* **12**(4), 143–144.
- Elston-Güttler, K. E., and Gunter, T. C. (2009). "Fine-tuned: Phonology and semantics affect first-to second-language zooming in," *J. Cogn. Neurosci.* **21**(1), 180–196.
- Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," *J. Acoust. Soc. Am.* **116**(4), 2365–2373.
- Francis, A. L., Tigchelaar, L. J., Zhang, R., and Zekveld, A. A. (2018). "Effects of second language proficiency and linguistic uncertainty on recognition of speech in native and nonnative competing speech," *J. Speech Lang. Hear. Res.* **61**(7), 1815–1830.
- Grynpas, J., Baker, R., and Hazan, V. (2011). "Clear speech strategies and speech perception in adverse listening conditions," paper presented at the International Congress of Phonetic Science.
- Hahne, A. (2001). "What's different in second-language processing? Evidence from event-related brain potentials," *J. Psycholing. Res.* **30**(3), 251–266.
- Hazan, V., Grynpas, J., and Baker, R. (2012). "Is clear speech tailored to counter the effect of specific adverse listening conditions?," *J. Acoust. Soc. Am.* **132**(5), EL371–EL377.
- IELTS (2017). <https://www.ielts.org/> (Last viewed June 1, 2020).
- Just, M. A., and Carpenter, P. A. (1993). "The intensity dimension of thought: Pupillometric indices of sentence processing," *Can. J. Exp. Psychol.* **47**(2), 310.
- Koch, X., and Janse, E. (2016). "Speech rate effects on the processing of conversational speech across the adult life span," *J. Acoust. Soc. Am.* **139**(4), 1618–1636.
- Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., and Skarzynski, H. (2012). "Processing load during listening: The influence of task characteristics on the pupil response," *Lang. Cogn. Proc.* **28**(4), 426–442.
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr., Cute, S. L., Humes, L. E., Dubno, J. R., and Eckert, M. A. (2013). "Pupil size varies with word listening and response selection difficulty in older adults with hearing loss," *Psychophysiology* **50**(1), 23–34.
- Kuipers, J. R., and Thierry, G. (2011). "N400 amplitude reduction correlates with an increase in pupil size," *Front. Human Neurosci.* **5**, 61.
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). "A cortical network for semantics:(de) constructing the N400," *Nat. Rev. Neurosci.* **9**(12), 920.
- Lecumberri, M. L. G., Cooke, M., and Cutler, A. (2010). "Non-native speech perception in adverse conditions: A review," *Speech Commun.* **52**(11), 864–886.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**(2B), 467–477.
- Li, P., Zhang, F., Tsai, E., and Puls, B. (2014). "Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool," *Bilingualism: Lang. Cogn.* **17**(03), 673–680.
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). "The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," *J. Speech Lang. Hear. Res.* **50**(4), 940–967.
- Mattys, S. L., Carroll, L. M., Li, C. K., and Chan, S. L. (2010). "Effects of energetic and informational masking on speech segmentation by native and non-native speakers," *Speech Commun.* **52**(11), 887–899.
- Mayo, L. H., Florentine, M., and Buus, S. (1997). "Age of second-language acquisition and perception of speech in noise," *J. Speech Lang. Hear. Res.* **40**(3), 686–693.
- Payton, K. L., Uchanski, R. M., and Braid, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **95**(3), 1581–1592.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., and Naylor, G. (2016). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)," *Ear and Hearing* **37**, 5S–27S.
- Piquado, T., Isaacowitz, D., and Wingfield, A. (2010). "Pupillometry as a measure of cognitive effort in younger and older adults," *Psychophysiology* **47**(3), 560–569.
- R Core Team. (2017). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/> (Last viewed June 1, 2020).
- Richer, F., and Beatty, J. (1985). "Pupillary dilations in movement preparation and execution," *Psychophysiology* **22**(2), 204–207.
- Rönnerberg, J., Holmer, E., and Rudner, M. (2019). "Cognitive hearing science and ease of language understanding," *Int. J. Audiology* **58**(5), 247–261.
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., and Rudner, M. (2013). "The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances," *Front. System Neurosci.* **7**, 31.
- Rönnerberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). "Cognition counts: A working memory system for ease of language understanding (ELU)," *Int. J. Audiology* **47**(sup2), S99–S105.
- Schluroff, M. (1982). "Pupil responses to grammatical complexity of sentences," *Brain Lang.* **17**(1), 133–145.
- Schmidtke, J. (2014). "Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry," *Front. Psych.* **5**, 137.
- Simantiraki, O., Cooke, M., and King, S. (2018). "Impact of different speech types on listening effort," *Proc. Interspeech* **2018**, 2267–2271.
- Smiljanić, R., and Bradlow, A. R. (2009). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," *Lang. Ling. Comp.* **3**(1), 236–264.
- Smiljanić, R., and Bradlow, A. R. (2011). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness," *J. Acoust. Soc. Am.* **130**(6), 4020–4031.
- Song, J., and Iverson, P. (2018). "Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents," *Cognition* **179**, 163–170.
- SRResearch (2009). "EyeLink 1000 User Manual Tower," SR Research, 1.5.0 edition, <http://sr-research.jp/support/EyeLink%201000%20User%20Manual%201.5.0.pdf> (Last viewed June 1, 2020).
- Tabachnick, B. G., Fidell, L. S., and Ullman, J. B. (2007). *Using Multivariate Statistics* (Pearson, Boston, MA), Vol. 5.
- Takata, Y., and Nábelek, A. K. (1990). "English consonant recognition in noise and in reverberation by Japanese and American listeners," *J. Acoust. Soc. Am.* **88**(2), 663–666.
- Van Engen, K. J., Phelps, J. E., Smiljanic, R., and Chandrasekaran, B. (2014). "Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker," *J. Speech Lang. Hear. Res.* **57**(5), 1908–1918.
- Wendt, D., Dau, T., and Hjortkjær, J. (2016). "Impact of background noise and sentence complexity on processing demands during sentence comprehension," *Front. Psychol.* **7**, 345.
- Wingfield, A., and Tun, P. A. (2007). "Cognitive supports and cognitive constraints on comprehension of spoken language," *J. Am. Acad. Audiol.* **18**(7), 548–558.
- Winn, M. B. (2016). "Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants," *Trends Hear.* **20**, 2331216516669723.
- Winn, M. B., Wendt, D., Koelewijn, T., and Kuchinsky, S. E. (2018). "Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started," *Trends Hear.* **22**, 2331216518800869.
- Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). "The pupil dilation response to auditory stimuli: Current state of knowledge," *Trends Hear.* **22**, 2331216518777174.
- Zekveld, A. A., and Kramer, S. E. (2014). "Cognitive processing load across a wide range of listening conditions: Insights from pupillometry," *Psychophysiology* **51**(3), 277–284.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," *Ear Hear.* **31**(4), 480–490.
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., Festen, J. M., Van Beek, J. H., and Rönnerberg, J. (2011). "The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise," *Ear Hear.* **32**(6), e16–e25.