Cognitive load elevates discrimination thresholds of duration, intensity, and f_0 for a synthesized vowel

Faith Chiu,[a] Lyndon L. Rakusen, and Sven L. Mattys
Department of Psychology, University of York, York, YO10 5DD, United Kingdom

(Received 3 May 2019; revised 25 June 2019; accepted 15 July 2019; published online 12 August 2019)

Dual-tasking negatively impacts on speech perception by raising cognitive load (CL). Previous research has shown that CL increases reliance on lexical knowledge and decreases reliance on phonetic detail. Less is known about the effect of CL on the perception of acoustic dimensions below the phonetic level. This study tested the effect of CL on the ability to discriminate differences in duration, intensity, and fundamental frequency of a synthesized vowel. A psychophysical adaptive procedure was used to obtain just noticeable differences (JNDs) on each dimension under load and no load. Load was imposed by *N*-back tasks at two levels of difficulty (one-back, two-back) and under two types of load (images, nonwords). Compared to a control condition with no CL, all *N*-back conditions increased JNDs across the three dimensions. JNDs were also higher under two-back than one-back load. Nonword load was marginally more detrimental than image load for intensity and fundamental frequency discrimination. Overall, the decreased auditory acuity demonstrates that the effect of CL on the listening experience can be traced to distortions in the perception of core auditory dimensions. © *2019 Acoustical Society of America*.
https://doi.org/10.1121/1.5120404

[DB] Pages: 1077–1084

I. INTRODUCTION

The growing interest in studying speech perception in realistic environments has led to an abundance of research conducted in sub-optimal listening conditions (for a review, see Mattys *et al.*, 2012). Conditions that affect the integrity of the speech signal (e.g., noise) have been investigated using primarily the concepts of energetic and informational masking (e.g., Brungart, 2001; Durlach *et al.*, 2003). Less is known about conditions that do not create acoustic interference to the speech signal itself, but place demands on cognitive functions and deplete processing resources necessary for speech perception (e.g., the effect of monitoring road traffic on the ability to perceive and understand speech).

Prior research has demonstrated effects of cognitive load (CL) on various levels of speech processing. At the sentence level, speech intelligibility in quiet or noise drops under divided attention (Best *et al.*, 2010). At sub-sentential levels, CL also modifies processing behavior. Whenever a lexically viable option is compared against a nonword, participants default to the lexical option under CL, downplaying the contribution of acoustic detail. For instance, in Mattys *et al.* (2009), participants had to decide which of two words—*mild* or *mile*—they heard at the beginning of a phrase like /maɪldɒpʃən/. The phrase varied in its realization, from /maɪld#ɒpʃən/ (mild#option) to /maɪl#dɒpʃən/ (mile#doption), via the manipulation of local coarticulation and word-boundary cues. Under CL, there was an increase in reporting the word that led to the lexically viable segmentation outcome (*mild*, leading to mild#option).

Evidence for a lexical drift was also found in phoneme identification tasks. Mattys and Wiget (2011) constructed a continuum between the word "gift" and the nonword "kift," varying the voice onset time (VOT) of the onset consonant in several steps from 0 ms to 48 ms. When asked to report whether they heard /g/ or /k/ at the beginning of those syllables, participants under CL reported /g/ more often, that is, the sound compatible with the word—as opposed to nonword—endpoint. Conversely, they reported more /k/ on a continuum from "kiss" to "giss," showing again lexically biased phoneme identification under CL. Thus, in tasks allowing access to lexical representations, participants under CL often default to lexically meaningful interpretations despite inconsistent acoustic cues. This may indicate lessened access to or integration of acoustic information under CL.

At even lower levels of speech processing, CL interferes with the listener's ability to make fine acoustic-phonetic judgements. For example, listeners showed reduced discriminability between syllables along a /gɪ-kɪ/ continuum when they simultaneously engaged in a visual search task (Mattys and Wiget, 2011) or a face recognition memory task (Mitterer and Mattys, 2017). Likewise, Mattys *et al.* (2014) noted that phoneme restoration, the illusion of hearing a phoneme when there is only noise, increased linearly as a function of the secondary task. Participants' ability to discriminate between a noise-overlaid phoneme and noise alone decreased under CL. This reveals that participants perceived and represented the incoming signal with less acoustic precision under load. Importantly, the effect was not affected by whether the phoneme-carrying stimuli were words or nonwords. This led the authors to conclude that CL interferes with early, low-level speech perception processes.

A central question for this study is how early in the perceptual system the effect of CL can be traced. There is some evidence that CL might alter the perception of core auditory dimensions. For instance, Casini *et al.* (2009) found a consistent underestimation of speech stimulus duration under CL.

[a]Electronic mail: faith.chiu@york.ac.uk

In their experiment, carried out in French, participants had to judge which of two spoken words they heard while performing a color-identification task as CL. The two words differed primarily in the duration of their vowel. In French, vowel length preceding a final obstruent is a cue to the voicing status of that obstruent with a short vowel cueing a voiceless obstruent (e.g., /ʃ/) and a longer vowel cueing a voiced one (e.g., /ʒ/). This phenomenon is known as pre-fortis clipping (Kingston and Diehl, 1994). When presented with stimuli along a /kaʃ-kaʒ/ continuum, participants reported more instances of the voiceless French word *cache* (short vowel) than the voiced word *cage* (long vowel) under CL than no CL. The authors interpreted their results as evidence that divided attention leads to a reduction in the perceived duration of sounds. This interpretation is rooted in models proposing internal clocks and prospective duration estimation (e.g., Treisman, 1963). In these models, input pulses (or samples) accumulate during a time-bound event, and the duration of the event is estimated as the tallied duration of the accumulated pulses (e.g., Gibbon *et al.*, 1984; Block *et al.*, 2010, for review). Under divided attention, where attention is intermittently diverted to a competing task, pulses are missed (Block and Zakay, 1996; Zakay and Block, 1995), causing duration to be underestimated.

The possibility that pulse-skipping is the general mechanism by which CL disrupts sound perception is challenged on several counts. First, CL-induced time-shrinkage effects are not consistently found for all duration-based speech cues. For example, using VOT as a cue to voicing (short, /g/; long, /k/), Mattys and Wiget (2011) did not find a tendency to report more /g/ than /k/ sounds under CL. Second, if CL led to pulse-skipping and reduced perceived duration, CL would be more likely to affect the perception of time-dependent (e.g., phoneme duration) than time-independent dimensions of speech (e.g., intensity and pitch). However, there is evidence that CL can affect non-durational dimensions. For example, Macdonald and Lavie (2011) showed that participants failed to notice the presence of brief near-threshold pure tones when they were simultaneously performing a visual discrimination task. This was replicated at the neuro-functional level by Molloy *et al.* (2015). Thus, there is some evidence that CL can affect not only duration perception but also intensity perception.

Despite these challenges, the hypothesis that pulse-skipping is the mechanism by which CL disrupts sound perception could be retained if we consider the broader implications of the claim. Psychoacoustic research on temporal integration shows that the precision with which one can judge the intensity and pitch of a tone is a function of the duration of the presented stimulus, with longer presentations yielding higher precision (e.g., Florentine, 1986; Florentine *et al.*, 1988; Moore, 1973; Plack and Carlyon, 1995). Improvement at longer durations is thought to result from the ability to compare and integrate more input samples and, hence, reduce error variance (e.g., Viemeister and Wakefield, 1991). If CL reduces the number of samples a listener is able to process, as predicted by the pulse-skipping hypothesis, CL should not only directly affect duration judgement but also indirectly affect intensity and pitch judgements.

To test the above hypotheses within a single study, we investigated the effect of CL on the discrimination of synthesized vowel-like sounds (vocoids) manipulated in duration, intensity, and f_0. These manipulations correspond to three basic perceptual dimensions of duration, loudness, and pitch. For each dimension, we used a psychophysical adaptive procedure to determine the just noticeable difference (JND) under no CL, perceptual load, low CL, and high CL. CL consisted of a visual one-back task (low CL) or two-back task (high CL). The perceptual load condition, which involved presenting the *N*-back stimuli without the *N*-back task itself, served as a baseline to delineate effects of perceptual load (selective attention) and CL (divided attention). The contrast between low and high CL directly assessed the effect of CL within divided attention. Furthermore, to test the generalizability of the CL effect, we contrasted CL that requires visual-only encoding (*N*-back task using images) and CL that requires subvocal auditory encoding (*N*-back rhyme task using nonwords; see Fig. 1 for an example).

If CL affects basic perceptual processes, it should elevate JNDs (i.e., poorer discrimination) for at least some of the dimensions tested. Of particular interest is whether CL will increase only duration JNDs, as predicted by a strict interpretation of the pulse-skipping hypothesis, or whether CL will increase not only duration but also intensity and f_0 JNDs, as predicted by a generalized version of the pulse-skipping hypothesis. The contrast between image CL and nonword CL will reveal whether the interference of CL on auditory perception depends on the representational format of CL. If it does, we expect that CL involving auditory representations (nonwords) will be more detrimental than CL involving visual representations (images).

II. METHOD

A. Participants

Ninety-six York-based university students participated in this study. Participants were randomly assigned to one of three groups: duration ($n = 32$, with 22 female; M_{age}, 21; range, 18–36); intensity ($n = 32$, with 21 female; M_{age}, 20.84; range, 18–35), and f_0 ($n = 32$, with 27 female; M_{age}, 19.78; range, 18–27). Participants were assessed for their hearing using pure tone audiometry in accordance to the 2011 British Society of Audiology recommended procedure. However, to keep testing time within reasonable limits, only 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz were tested. None of the participants exceeded a threshold of 20 dB hearing level (HL) at any of the four frequencies in either ear.

All participants reported normal or corrected-to-normal vision, and none reported any speech and/or hearing impairments. All participants were native English speakers. Relevant to methodological considerations described later, none of them reported any proficiency of Chinese, Tamil, Bengali, or Gujarati, or had parents, family members, or partners with Chinese, Tamil, Bengali, or Gujarati knowledge. All participants gave informed consent and were

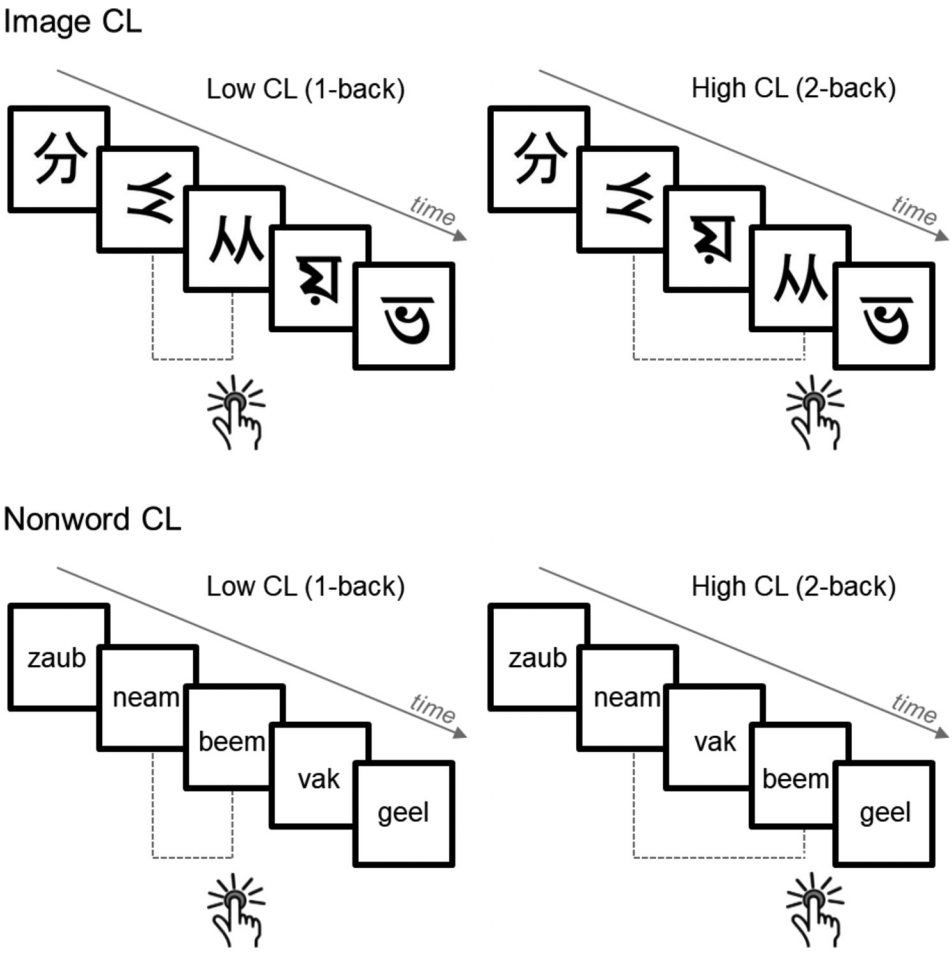

FIG. 1. Illustration of the CL task for images (top) and nonwords (bottom).

compensated with monetary payment or course credit. The study was approved by the University of York Departmental Ethics Committee (identification number 2018-712).

B. Materials

1. Auditory stimuli for the JND task

To obtain JNDs for duration, intensity, and f_0, 181 audio files were synthesized. These consisted of a base stimulus from which each of the 3 acoustic dimensions deviated in 60 steps. To create the base stimulus, a male monolingual British English speaker was recorded in a sound-attenuated booth producing several instances of the vowel /ɑ/ with a flat pitch contour. Recordings were made with a dynamic cardioid microphone (SHURE SM58, Niles, IL) through an audio interface recorder (USB Dual Pre, Applied Research and Technology, Ontario, Canada) using Praat software (Boersma and Weenink, 2018) at a sampling rate of 44.1 kHz. The best exemplar, as judged by a trained phonetician, was selected to provide reference formant values for the creation of the base stimulus and the deviant stimuli. The base stimulus was a Klatt-synthesized (Klatt and Klatt, 1990) vowel-like steady-state token with the following parameters: $F1$ 836 Hz, $F2$ 1152 Hz, $F3$ 2741 Hz, f_0 150 Hz, 500 ms, 60 dB sound pressure level (SPL). Deviant stimuli differed from the base stimulus in one dimension only (duration, intensity, or f_0) with all other acoustic parameters kept identical to the base. Deviant duration stimuli ranged from 520 to 800 ms in 60 steps of 5 ms each. Deviant intensity stimuli ranged from $60\frac{1}{6}$ to 70 dB SPL in 60 steps of $\frac{1}{6}$ dB each. Deviant f_0 stimuli ranged from 140.05 to 153.00 Hz in 60 steps of 0.05 Hz each. 20-ms on and off ramps were applied to all sounds. The presentation levels of the stimuli were established using a Brüel and Kjær (B&K, Nærum, Denmark) 4153 artificial ear, a B&K 4189 $\frac{1}{2}$ inch microphone, a B&K 4767 preamplifier, and a B&K 2260 sound level meter.

2. Visual stimuli for the CL tasks

Two types of visual stimuli were used as CL: images and written nonwords. All stimuli were in black against a white background. Image stimuli consisted of 27 4-stroke Chinese characters and 27 characters drawn from a mixture of Bengali, Gujarati, and Tamil characters selected from Gennari *et al.* (2018). Based on our participant selection criteria, all characters were deemed unnameable and therefore only encodable visually.

Nonword stimuli were 54 written monosyllabic stimuli modified from a combination of nonwords from Palmer and Mattys (2016) and McQueen (1993). Nonword structure was (C)CVC(C), where C stands for consonant, and V, vowel, with optional segments in parentheses. Rhyming nonwords for the one-back and two-back rhyme CL task included orthographically similar (e.g., *dird*, *chird*) and orthographically dissimilar but phonologically similar (e.g., *dird*, *vurd*) nonwords. The mixture of the two types of stimuli was meant to encourage phonological processing during the

N-back tasks so that participants could not merely rely on orthography to complete the task.

C. Design and procedure

Dimension (duration, intensity, f_0) was a between-participant factor. For each dimension, JNDs were obtained under perceptual load, low CL (one-back), and high CL (two-back). Each of these three conditions was presented with one of two load types: images or nonwords. In the perceptual load condition, the image and nonword stimuli were meant to provide a sensory input equivalent to that of the CL conditions, but no *N*-back task was required from them.

Participants performed all tasks in a sound-attenuated booth. Data were gathered over two sessions, which were separated by at least one day. In session 1, participants completed a pure tone audiometry test, the JND task under no load (audio-only), the JND task under perceptual load, and four short CL-only blocks (low CL image, low CL nonword, high CL image, high CL nonword). The audio-only condition served as practice for the JND task. The perceptual load condition, which served as a baseline for the CL conditions in session 2, was administered before the CL conditions to ensure that participants remained naive to the purpose of the visual stimuli. The CL-only blocks allowed participants to familiarise themselves with the CL tasks used in session 2.

In session 2, participants did the JND task under low CL image, low CL nonword, high CL image, and high CL nonword. The JND obtained for each of these four conditions was the average of two paired adaptive tracks, one with the deviant corresponding to the larger value on the manipulated acoustic dimension (longer, louder, higher in pitch) and the other with the deviant corresponding to the smaller value on that dimension (shorter, softer, lower in pitch). The order of the eight tracks (2 CL levels × 2 load types × 2 paired tracks) was counterbalanced across sets of eight participants.

1. Auditory discrimination task for JNDs

JNDs were estimated based on a three-interval two-alternative forced-choice task (3I-2AFC). On each trial, participants heard three consecutive auditory stimuli. Using the "S" and "D" keys on a computer keyboard, they had to decide which of the second or third stimulus was the deviant on the relevant dimension. Depending on which of the paired tracks was played, the deviant could be longer or shorter (duration), louder or softer (intensity), or lower or higher in pitch (f_0). Thus, the base stimulus was the stimulus corresponding to either the first step on the continuum of the relevant dimension or the last one. Before each track, participants were informed of the nature of the deviance they had to listen for. The inter-stimulus interval was 500 ms. Participants could only respond after all three stimuli were heard.

A two-down/one-up adaptive track establishing the 70.7% discrimination threshold (Levitt, 1971) started with the deviant stimulus as the furthest step from the standard stimulus (e.g., step 60 relative to step 0). For example, the participant in the duration condition performing the longer-deviant track started with a 500-ms stimulus as the base and a 800-ms stimulus as the deviant (500-500-800 or 500-800-500). As the task progressed, the duration of the deviant was reduced as a function of the participant's response (500-500-750 or 500-750-500). Step size corresponded to ten units of the continuum until the first reversal (50 ms, or 1.6 dB, or 0.5 Hz), decreasing over the first three reversals to one unit (5 ms, or 0.16 dB, or 0.05 Hz). The task ended after 16 reversals or a maximum of 70 trials. The JND was estimated by taking the mean value of the final 8 reversals or, if 70 trials were necessary, the mean of all the reversals after the minimum step size had been reached. A similar progression applied to the paired track, with the 800-ms stimulus being the standard and the 500-ms stimulus being the deviant.

2. Low CL and high CL tasks

Figure 1 illustrates the procedure for the CL one-back and two-back tasks. Visual stimuli (images or nonwords) for the *N*-back tasks were displayed for 750 ms each with a 250 ms inter-stimulus interval (a white screen). In the low- and high-CL conditions, participants engaged in a one-back and a two-back task, respectively. Similar to the procedure described in Palmer and Mattys (2016), in the image condition, participants were instructed to press a key with their non-dominant hand whenever they saw an image that matched the one immediately preceding it (one-back) or the image presented two images before (two-back). Repeated images appeared either in an identical orientation or were left-rotated by 90 degrees. In the nonword condition, images were replaced with pronounceable nonwords. Participants pressed a key whenever they saw a nonword that rhymed with the one immediately preceding it (one-back) or the nonword presented two nonwords before (two-back). After each image repetition or rhyming nonword, there was a range of 2–4 intervening stimuli before the next repetition. The stream of visual stimuli stopped at the end of the JND track. Therefore, the total number of repetitions varied from track to track. Participants were instructed to try and perform both tasks equally well. Given the inter-tone-interval depended on the participant's response time, there was no systematic alignment between tones and visual stimuli.

In the perceptual load condition, participants were shown the same images or nonwords as in the CL conditions, but there were no one-back or two-back repetitions. Participants were instructed to pay attention to the visual stimuli but were not given an active task. In the visual-only condition (no auditory task), participants performed the 1-back and 2-back tasks on a total of 30 pairs of matching stimuli (image or nonword).

III. RESULTS

This section describes two sets of analyses: (1) JNDs across load conditions and (2) performance on the CL task and its relation to JNDs. Figure 2 displays the results in both tasks.

A. JNDs

1. Duration

Average JNDs for the audio-only, perceptual load, low CL, and high CL conditions are shown in Fig. 2.[1] We first aimed to answer whether there was a detrimental effect

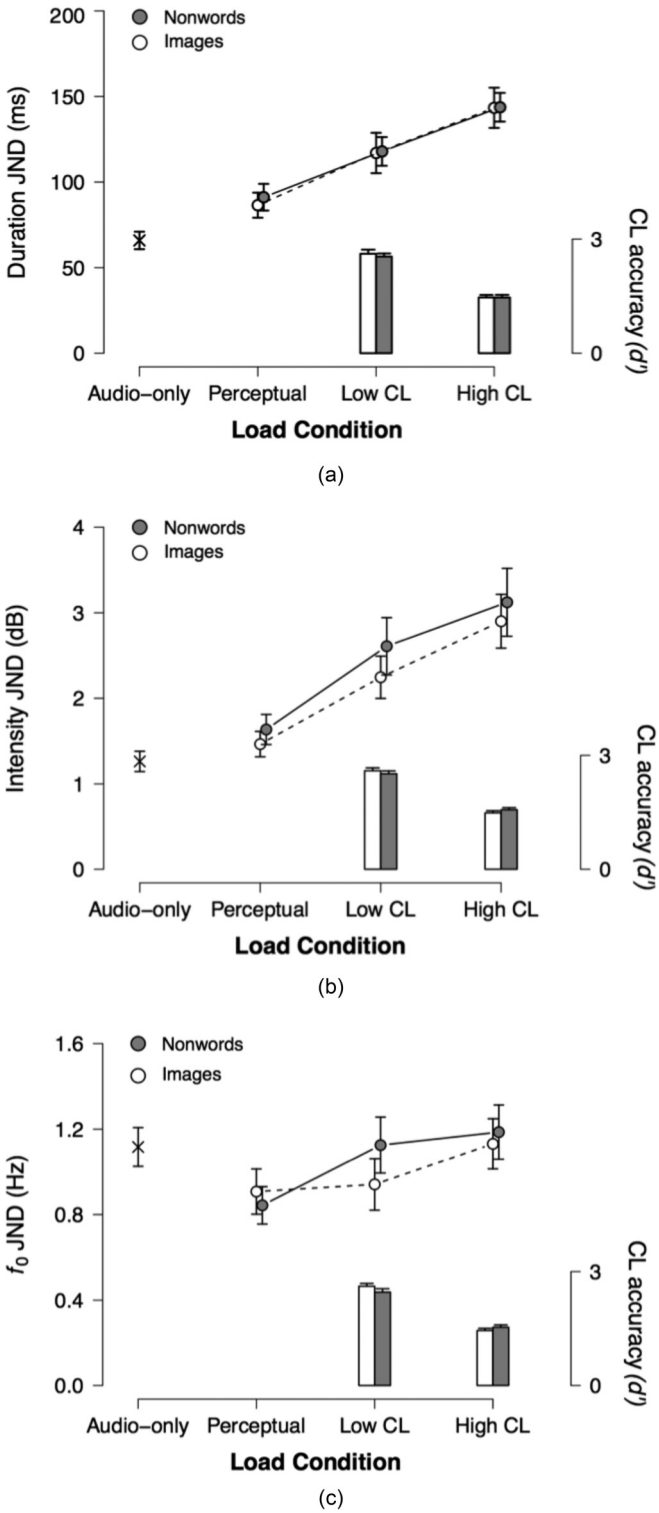

FIG. 2. JNDs in the audio-only, perceptual load (no secondary task), low CL (one-back task), and high CL (two-back task) conditions for (a) duration, (b) intensity, and (c) f_0. Load type (images vs nonwords) is shown as separate lines. Error bars indicate the standard error of the mean. The bar chart, scaled on the right-hand y axis, shows CL performance (d') as a function of CL level (low vs high) and load type (images vs nonwords).

caused by CL (averaged across low and high) on JNDs relative to the baseline perceptual load condition, a contrast essentially pitting divided against selective attention. An analysis of variance (ANOVA) of JNDs across attention type (selective vs divided attention) and load type (image vs nonword) indicated a main effect of attention type, $F(1,31) = 41.11$, $p < 0.001$, $\eta p^2 = 0.570$ with higher JNDs under divided attention [$M = 131$ ms, standard deviation (SD) = 51 ms] than selective attention ($M = 89$ ms, SD = 39 ms). Neither load type, $F(1,31) < 1$, nor the interaction term, $F(1,31) < 1$, was significant. Thus, the smallest perceptible difference in duration increased when participants had to divide their attention between the auditory task and a visual CL task, and it did so regardless of the type of encoding required by the CL task.

We then tested whether the magnitude of CL had a detrimental effect on JNDs within divided attention. An ANOVA with CL level (low vs high) and load type (image vs nonword) showed a main effect of CL level, $F(1,31) = 16.56$, $p < 0.001$, $\eta p^2 = 0.348$, with higher JNDs under high CL ($M = 144$ ms, SD = 57 ms) than low CL ($M = 117$ ms, SD = 52 ms). There was no effect of load type, $F(1,31) < 1$, or interaction, $F(1,31) < 1$.

In sum, duration discrimination was worsened not only by divided attention, but also by the difficulty of the secondary task within divided attention. Whether CL required auditory or visual encoding was inconsequential.

2. Intensity

Similar to duration JNDs, intensity JNDs were higher under divided attention ($M = 2.72$ dB, SD = 1.65 dB) than under selective attention ($M = 1.55$ dB, SD = 0.84 dB), $F(1,31) = 25.20$, $p < 0.001$, $\eta p^2 = 0.448$. However, intensity JNDs were also affected by load type, $F(1,31) = 6.67$ $p = 0.015$, $\eta p^2 = 0.183$, with higher JNDs under nonword CL ($M = 2.25$ dB, SD = 1.25 dB) than image CL ($M = 2.02$ dB, SD = 1.06 dB). The interaction term was not significant, $F(1,31) < 1$.

A comparison between low and high CL revealed a main effect of CL level, $F(1,31) = 9.89$, $p = 0.004$, $\eta p^2 = 0.242$, with higher JNDs under high CL ($M = 3.01$ dB, SD = 1.86 dB) than low CL ($M = 2.43$ dB, SD = 1.59 dB). There was also a marginal effect of load type, $F(1,31) = 3.23$, $p = 0.082$, $\eta p^2 = 0.094$, showing higher JNDs under nonword CL ($M = 2.87$ dB, SD = 1.91 dB) than image CL ($M = 2.57$ dB, SD = 1.49 dB). The interaction term was not significant, $F(1,31) < 1$.

Thus, intensity discrimination was worse under divided attention than selective attention and, within divided attention, worse under high CL than low CL. This pattern is similar to that for duration JNDs. Unlike duration JNDs, however, intensity JNDs were higher when CL involved auditory encoding (nonword CL task) than visual encoding (image CL tasks).

3. f_0

JNDs for f_0 were higher under divided attention ($M = 1.10$ Hz, SD = 0.66 Hz) than selective attention ($M = 0.88$ Hz, SD = 0.53 Hz), $F(1,31) = 11.13$, $p = 0.002$, $\eta p^2 = 0.264$. The effect of load type was not significant, $F(1,31) < 1$. An interaction between attention type and load type, $F(1,31) = 5.51$, $p = 0.02$, $\eta p^2 = 0.151$, showed that the effect of attention type was significant under nonword load, $F(1,31) = 14.52$, $p < 0.001$, $\eta p^2 = 0.319$, and there was also a marginal effect under image load, $F(1,31) = 3.26$, $p = 0.08$, $\eta p^2 = 0.095$.

A comparison between low and high CL revealed a main effect of CL level, $F(1,31) = 7.15$, $p = 0.012$, $\eta p^2 = 0.187$, with higher JNDs under high CL ($M = 1.16$ Hz, SD = 0.67 Hz) than low CL ($M = 1.03$ Hz, SD = 0.69 Hz). There was also a significant effect of load type, $F(1,31) = 6.60$, $p = 0.015$, $\eta p^2 = 0.176$, showing higher JNDs under nonword CL ($M = 1.16$ Hz, SD = 0.70 Hz) than image CL ($M = 1.04$ Hz, SD = 0.65 Hz). The interaction term was not significant, $F(1,31) = 1.83$, $p = 0.18$.

In sum, although less pronounced than those for intensity, the f_0 JND patterns broadly aligned with intensity in showing a detrimental effect of divided attention and CL level on f_0 discrimination, and greater interference from a load requiring auditory encoding than visual encoding.

B. CL task performance

Performance on the visual task was measured as the ability to discriminate between repeated images (or rhyming nonwords) and non-repeated images (or non-rhyming nonwords), using the discriminability index d' from signal detection theory (Green and Swets, 1966). For each condition, the hit rate was calculated as the number of correct responses to repeated/rhyming stimuli (one-back or two-back) divided by the total number of repetitions/rhymes encountered during that condition. The false alarm rate was calculated as the total number of incorrect responses to non-repeated/rhyming stimuli divided by the total number of non-repeated/rhyming stimuli. Results are plotted as bars in Fig. 2.

An ANOVA with CL level (low vs high), load type (image vs nonword), and dimension (duration, intensity, f_0) performed on d' values showed a main effect of CL level, with higher d' for low CL than high CL, $F(2,93) = 1018.69$, $p < 0.001$, $\eta p^2 = 0.92$, which confirms that the one-back task was less demanding than the two-back task. However, the CL level effect interacted with load type, $F(2,93) = 9.06$, $p = 0.003$, $\eta p^2 = 0.090$: The difference between low and high CL was more pronounced in the image than the nonword condition. None of the other main effects or interactions reached significance. In particular, the lack of a load type effect, $F(2,93) < 1$, suggested that the image and nonword CL tasks were broadly comparable in complexity despite involving different encoding modalities.

C. Relation between JNDs and CL performance

Potential links between auditory discrimination (JNDs) and performance on the CL task (d') were assessed via Pearson's correlation coefficients with corrections for multiple comparisons (Bonferroni) where appropriate. Of interest was whether performance on the auditory task traded off with performance on the CL task. JND values used in these tests were absolute values without any subtraction of JNDs from perceptual load. Correlations were first calculated for each individual condition of the load level × load type × dimension design (12 correlations in all). Within each dimension, additional correlations were calculated with data collapsed across either load level or load type, or across both. None of these correlations reached significance (all $p > 0.05$). In fact, a majority of them showed the opposite valence, indicating that, if at all, participants who were better at auditory discrimination (lower JND) performed better on the concurrent CL task (higher d'). This result may be indicative of the lack of resource-sharing between the two tasks, at least within this particular group of participants.

We then investigated the link between JND and d' in terms of CL cost. We asked whether the cost of performing the auditory task under CL was correlated with performance in the CL task. JND cost was measured as the JND of each CL condition of the design minus the JND from the relevant perceptual load condition. JND cost was then correlated with the corresponding d'. None of the correlations reached significance. Thus, there was no trade-off between the cost of performing the auditory task under CL and the ability to perform the CL task. Finally, we correlated the JND difference between the low and high CL conditions with the d' difference between the low and high CL conditions. Of interest was whether listeners who showed a large d' difference between the low and high CL conditions would also show a large JND difference between those two conditions. Such a correlation would indicate that the added difficulty of performing the two-back task compared to the one-back task would be mirrored by a corresponding increase in JNDs. Again, none of the correlations reached significance.

IV. DISCUSSION

Evidence has shown that CL disrupts various components of the listening experience. In this study, we aimed to find out if CL can also alter basic auditory processes below the level of the phoneme. We measured the impact of CL on the discrimination of synthesized vowels manipulated in terms of their duration, intensity, or f_0. For each dimension, JNDs were estimated under three load levels (perceptual, low CL, high CL) and two load types (*N*-back task on images vs *N*-back rhyme task on nonwords). Of particular interest was whether the detrimental effect of CL on JNDs, if present, was limited to the duration dimension, as per a strict interpretation of the pulse-skipping hypothesis (Block and Zakay, 1996; Burle and Casini, 2001; Casini and Macar, 1997; Casini *et al.*, 2009; Zakay and Block, 1995), or whether it generalized across all three dimensions, as might be predicted by a failure of temporal integration under CL (Florentine, 1986; Moore, 1973; Viemeister and Wakefield, 1991).

We found that JNDs increased under divided attention (averaged across low and high CL) compared to selective attention (perceptual load) and under high CL compared to low CL. Importantly, CL affected JNDs in all tested dimensions, not only duration. This is inconsistent with a strict interpretation of the pulse-skipping hypothesis, which posits that a loss of input samples during divided attention should affect primarily duration judgement. Therefore, we propose that a loss of pulses disrupts the ability to judge not only duration but also the detailed content of the remaining pulses (e.g., intensity envelope and spectral information). This possibility is supported by evidence that intensity and f_0 are more difficult to estimate for stimuli played for short than long durations (e.g., Florentine, 1986; Florentine *et al.*, 1988; Moore, 1973; Plack and Carlyon, 1995; Viemeister

and Wakefield, 1991). Thus, increased JNDs under divided attention could be explained by the reduced number of auditory samples available in that condition. Similarly, increased JNDs under high than low CL could be explained by the larger number of attentional switches required by the high than low CL stimuli and the consequently smaller number of samples extracted from the auditory stimulus. We should note, however, that a caveat for the temporal integration interpretation is that the detrimental effect of short duration on intensity and f_0 estimation has been demonstrated mostly for durations under 200 ms (Plack and Carlyon, 1995), which is shorter than the duration of our stimuli. We must therefore assume that additional factors are likely to have contributed to the effect of CL on those two dimensions.

We also found some evidence of an effect of load type on intensity and f_0 judgements but not on duration. Intensity and f_0 JNDs were higher when the load task involved processing nonwords than images. We propose that the detrimental effect of CL is larger when its representational format competes with that of the auditory task. Since the CL rhyme task requires phonological encoding of the nonwords, such phonological representations would interfere with the processing of the auditory stimuli more than the stimuli in the image task. This is consistent with a domain-specific view of attentional allocation, which posits that attentional systems are dedicated to specific modalities (Adcock *et al.*, 2000; Petersen and Posner, 2012; Woodruff *et al.*, 1996). The same reasoning would hold for f_0 discrimination.

The fact that duration judgement was not affected by load type can be explained if we assume that, unlike intensity and pitch, the estimation of sound duration does not require the encoding of either envelope or spectral properties of the stimulus—detection of abrupt energy changes would suffice. The potential for interference from auditorily encoded CL would therefore be smaller in the case of duration estimation than intensity and pitch estimation. Interestingly, this possibility is in line with the claim by Casini *et al.* (2009) that the perceptual estimation of sound duration is governed by a general timing system and not by a sound- or language-specific system. Although speculative at this stage, it is possible that the effect of CL found in all three auditory dimensions of our design (duration, intensity, and f_0) involved one mechanism for duration discrimination and another for intensity and f_0. Duration discrimination could tap into a general timer system that drops content-free pulses under CL, whereas, for intensity and f_0 discrimination, the dropped pulses would contain acoustic information necessary for precise estimation of the auditory signal.

In sum, this study demonstrated that CL affects auditory perception at a lower level than has previously been shown, and it does so across several core auditory dimensions. CL effects were evident in duration, intensity, and f_0 discrimination with only the latter two displaying an additional cost under CL engaging auditory representations. The results can be interpreted within a broader version of the pulse-skipping hypothesis, in which loss of input samples under CL affects not only duration judgement but also the accuracy with which intensity and f_0 are encoded.

ACKNOWLEDGMENTS

This study was supported by a research grant from the Economic and Social Research Council (ESRC) to S.L.M. (Grant No. ES/R004722/1).

[1]Although included in Fig. 2, the audio-only condition was used primarily as practice for the JND task. In the duration and intensity dimensions, the JND for the audio-only condition was lower than the perceptual load condition, $t(31) = 2.34$, $p = 0.03$, and $t(31) = 3.78$, $p = 0.001$, respectively (averaged across load types). However, it was higher in the f_0 dimension, $t(31) = -3.54$, $p = 0.001$. The unexpected improvement of f_0 JND under perceptual load could be due to a particularly pronounced practice effect in that condition, since participants always completed the audio-only condition before the perceptual load condition. The reason why this practice effect would be greater for f_0 than duration and intensity is unclear. To test whether this contrast was robust or a random occurrence, we ran 24 participants on the audio-only and perceptual load conditions for duration, intensity, and f_0, counterbalancing the order of all conditions. As before, duration JNDs were lower in the audio-only condition ($M = 67$ ms, SD = 26 ms) than in the perceptual load condition, ($M = 82$ ms, SD = 35 ms), $t(24) = 2.17$, $p = 0.04$. The JND difference between audio-only and perceptual load was not significant in the intensity dimension ($M = 1.28$ dB, SD = 0.66 dB; $M = 1.30$ dB, SD = 0.52 dB; $t(24) = 0.20$, $p = 0.84$) or in the f_0 dimension ($M = 0.93$ Hz, SD = 0.60 Hz; $M = 0.97$ Hz, SD = 0.56 Hz; $t(24) = 0.34$, $p = 0.73$). These results indicate that the unexpected decrease in f_0 JNDs under perceptual load was probably a random occurrence related to the order of conditions.

Adcock, R. A., Constable, R. T., Gore, J. C., and Goldman-Rakic, P. S. (**2000**). "Functional neuroanatomy of executive processes involved in dual-task performance," Proc. Natl. Acad. Sci. **97**(7), 3567–3572.

Best, V., Gallun, F. J., Mason, C. R., Kidd, G., Jr., and Shinn-Cunningham, B. G. (**2010**). "The impact of noise and hearing loss on the processing of simultaneous sentences," Ear Hear. **31**(2), 213–220.

Block, R. A., Hancock, P. A., and Zakay, D. (**2010**). "How cognitive load affects duration judgments: A meta-analytic review," Acta Psychol. **134**(3), 330–343.

Block, R. A., and Zakay, D. (**1996**). "Models of psychological time revisited," Time Mind **33**, 171–195.

Boersma, P., and Weenink, D. (**2018**). "Praat: Doing phonetics by computer (version 6.0.43)[computer program]," http://www.praat.org (Last viewed May 3, 2019).

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**(3), 1101–1109.

Burle, B., and Casini, L. (**2001**). "Dissociation between activation and attention effects in time estimation: Implications for internal clock models," J. Exp. Psychol. **27**(1), 195–205.

Casini, L., Burle, B., and Nguyen, N. (**2009**). "Speech perception engages a general timer: Evidence from a divided attention word identification task," Cognition **112**(2), 318–322.

Casini, L., and Macar, F. (**1997**). "Effects of attention manipulation on judgments of duration and of intensity in the visual modality," Mem. Cognit. **25**(6), 812–818.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**(6), 2984–2987.

Florentine, M. (**1986**). "Level discrimination of tones as a function of duration," J. Acoust. Soc. Am. **79**(3), 792–798.

Florentine, M., Fastl, H., and Buus, S. R. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**(1), 195–203.

Gennari, S. P., Millman, R. E., Hymers, M., and Mattys, S. L. (**2018**). "Anterior paracingulate and cingulate cortex mediates the effects of cognitive load on speech sound discrimination," Neuroimage **178**, 735–743.

Gibbon, J., Church, R. M., and Meck, W. H. (**1984**). "Scalar timing in memory," Ann. N.Y. Acad. Sci. **423**(1), 52–77.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York), Vol. 1.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**(3), 419–454.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**(2), 820–857.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2B), 467–477.
Macdonald, J. S., and Lavie, N. (**2011**). "Visual perceptual load induces inattentional deafness," Atten., Percept., Psychophys. **73**(6), 1780–1789.
Mattys, S. L., Barden, K., and Samuel, A. G. (**2014**). "Extrinsic cognitive load impairs low-level speech perception," Psychon. Bull. Rev. **21**(3), 748–754.
Mattys, S. L., Brooks, J., and Cooke, M. (**2009**). "Recognizing speech under a processing load: Dissociating energetic from informational factors," Cognit. Psychol. **59**(3), 203–243.
Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (**2012**). "Speech recognition in adverse conditions: A review," Lang. Cognit. Process. **27**(7-8), 953–978.
Mattys, S. L., and Wiget, L. (**2011**). "Effects of cognitive load on speech recognition," J. Mem. Lang. **65**(2), 145–160.
McQueen, J. M. (**1993**). "Rhyme decisions to spoken words and nonwords," Mem. Cognit. **21**(2), 210–222.
Mitterer, H., and Mattys, S. L. (**2017**). "How does cognitive load influence speech perception? An encoding hypothesis," Atten., Percept., Psychophys. **79**(1), 344–351.
Molloy, K., Griffiths, T. D., Chait, M., and Lavie, N. (**2015**). "Inattentional deafness: Visual load leads to time-specific suppression of auditory evoked responses," J. Neurosci. **35**(49), 16046–16054.
Moore, B. C. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**(3), 610–619.
Palmer, S. D., and Mattys, S. L. (**2016**). "Speech segmentation by statistical learning is supported by domain-general processes within working memory," Q. J. Exp. Psychol. **69**(12), 2390–2401.
Petersen, S. E., and Posner, M. I. (**2012**). "The attention system of the human brain: 20 years after," Annu. Rev. Neurosci. **35**, 73–89.
Plack, C. J., and Carlyon, R. P. (**1995**). "Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics," J. Acoust. Soc. Am. **98**(3), 1355–1364.
Treisman, M. (**1963**). "Temporal discrimination and the indifference interval: Implications for a model of the 'internal clock,' " Psychol. Monogr. **77**, 1–31.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**(2), 858–865.
Woodruff, P. W., Benson, R. R., Bandettini, P. A., Kwong, K. K., Howard, R. J., Talavage, T., and Rosen, B. R. (**1996**). "Modulation of auditory and visual cortex by selective attention is modality-dependent," Neuroreport **7**, 1909–1913.
Zakay, D., and Block, R. A. (**1995**). "An attentional-gate model of prospective time estimation," in *Time and the Dynamic Control of Behavior*, edited by M. Richelle, V. DeKeyser, G. d'Ydewalle, and A. Vandierendock (Université de Liége, Liége), pp. 167–178.