**SUMMARY OF DATASET**

This dataset contains measures of just-noticeable differences, or JND, of duration (ms), intensity (dB), and pitch or F0 (Hz) for a synthetic vowel. JNDs were measured under 4 acoustic conditions: (1) Audio-only, i.e., with no secondary task, (2) Perceptual load, i.e., with concurrent visual stimuli but no requirement to pay attention to them, (3) Low cognitive load in the form of a 1-back task on visual stimuli, and (4) High cognitive load in the form of a 2-back task on visual stimuli. In addition, the 1-back and 2-back tasks involved either meaningless images or pronounceable nonwords. Performance on the 1-back and 2-back tasks, labelled CL accuracy, was also measured using the d' index from Signal Detection Theory.

Chiu, F., Rakusen, L. L., & Mattys, S. L. (2019). Cognitive load elevates discrimination thresholds of duration, intensity, and f 0 for a synthesized vowel. *The Journal of the Acoustical Society of America*, *146*(2), 1077-1084.

**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 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 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 three 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) through an audio interface recorder (USB Dual Pre) using Praat software (Boersma & 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-synthesised vowel-like steady-state token with the following parameters: *F*1 836 Hz, *F*2 1152 Hz, *F*3 2741 Hz, *f0* 150 Hz, 500 ms, 60 dB 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 601/6 to 70 dB SPL in 60 steps of 1/6 dB each. Deviant *f0* 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 & Kjær (B&K) 4153 artificial ear, a B&K 4189 ½ 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 four-stroke Chinese characters and 27 characters drawn from a mixture of Bengali, Gujarati, and Tamil characters selected from Gennari, Millman, Hymers, and Mattys (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), with optional segments in parentheses. Rhyming nonwords for the 1-back and 2-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 (1-back), and high CL (2-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 naïve 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 zero). For example, 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 eight 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 1-back and 2-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 1-back and a 2-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 (1-back) or the image presented two images before (2-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 (1-back) or the nonword presented two nonwords before (2-back). After each image repetition or rhyming nonword, there was a range of two to four 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 1-back or 2-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).

Figure 1

