Financial reward has differential effects on behavioural and self-report measures of listening effort

Peter J. Carolan¹,², Antje Heinrich¹,², Kevin J. Munro¹,² and Rebecca E. Millman¹,²

¹Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, UK; ²NIHR Manchester Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

ABSTRACT

Objectives: To investigate the effects of listening demands and motivation on listening effort (LE) in a novel speech recognition task.

Design: We manipulated listening demands and motivation using vocoded speech and financial reward, respectively, and measured task performance (correct response rate) and indices of LE (response times (RTs), subjective ratings of LE and likelihood of giving up). Effects of inter-individual differences in cognitive skills and personality on task performance and LE were also assessed within the context of the Cognitive Energetics Theory (CET).

Study sample: Twenty-four participants with normal-hearing (age range: 19 – 33 years, 6 male).

Results: High listening demands decreased the correct response rate and increased RTs, self-rated LE and self-rated likelihood of giving up. High financial reward increased subjective LE ratings only. Mixed-effects modelling showed small fixed effects for competitiveness on LE measured using RTs. Small fixed effects were found for cognitive skills (lexical decision RTs and backwards digit span) on LE measured using RTs and correct response rate, respectively.

Conclusions: The effects of listening demands on LE in the speech recognition task aligned with CET, whereas predictions regarding the influence of motivation, cognitive skills and personality were only partially supported.

Abbreviations: CET: Cognitive Energetics Theory; ELU: Ease of Language Understanding; FUEL: Framework for Understanding Effortful Listening; LE: Listening effort; MIT: Motivational Intensity Theory; MSE: Mean square error; NH: Normal hearing; RAU: Rationalised arcsine unit; RT: Response time; SNR: Signal-to-noise ratio

Introduction

Listening effort (LE) has been defined as ‘the mental exertion required to attend to, and understand, an auditory message’ (McGarrigle et al. 2014, p 434). A number of subjective (e.g. the NASA task load index (NASA-TLX; Hart and Staveland 1988), behavioural (e.g. RTs) and physiological (e.g. cardiac reactivity) measures of LE have been proposed (see McGarrigle et al. 2014 for a review). The Framework for Understanding Effortful Listening (FUEL), defines LE as the ‘deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a [listening] task’ (Pichora-Fuller et al. 2016, p. 10 S). According to this definition, the allocation of cognitive resources is not an automatic response to increased listening demands, but occurs only when a listener is motivated to achieve a particular goal.

The conceptual understanding of motivation outlined by FUEL builds upon Motivational Intensity Theory (MIT; Brehm and Self 1989), which posits that effort expenditure is influenced by both motivation and task demands. Most importantly, MIT predicts an interaction between listening demands and motivation. For relatively easy tasks, resource conservation limits the amount of effort expended, whereas predictions regarding the influence of motivation, cognitive skills and personality are only partially supported.

Nevertheless, there are inconsistencies both within and between studies that do not fit with MIT predictions based solely on task demands and motivation. When multiple LE measures are used within the same study, these do not always show consistent effects of listening demands and motivation, for example, a tone discrimination task resulted in changes in cardiac reactivity but no effects on performance accuracy or response times (RTs; Richter 2016). These differential effects may be due to the multi-dimensionality of LE, with different outcome measures tapping into interrelated aspects of the LE construct (McMahon et al. 2016; Strauss and Francis 2017; Hughes et al. 2018; Strand et al. 2018; Alhanbal © 2020; Herrmann and Johnsrude 2020). Moreover, additional factors may moderate the
relationship between task difficulty and motivation such as different ways of operationalising motivation (Picou and Ricketts 2014; c.f. Koelwijn et al. 2018; Richter 2016). For instance, no significant effect on RTs was found when motivating participants with financial reward (Richter 2016), but ‘threat of evaluation’ decreased RTs in an auditory oddball task (Carrillo de la Peña and Cadeveira 2000). Evaluative threat may increase arousal, a factor which has been demonstrated to result in faster RTs (Hackley and Valle-Inclán 1998); in other task designs increased LE may be reflected in slowed responses due to greater cognitive processing (Pisoni and Tash 1974; Marslen-Wilson and Tyler 1980). The slowing of RTs with increased listening demands has been interpreted as reflecting increased LE (Houben, van Doorn-Bierman, and Dreschler 2013). However, RTs are not a ‘process pure measure’ (Pichora-Fuller et al. 2016, 195); therefore, changes in RT may also reflect other aspects of cognition, such as memory.

One aspect that is not considered by FUEL/MIT but which may well influence motivation and effort expenditure is personality traits, such as the need for closure. Need for closure refers to an individual’s tendency to conserve resources. The balance between these forces is assumed to govern how much effort is exerted. CET makes a distinction between the maximum energy an individual is willing to mobilise to achieve a specific goal and the actual energy used, which depends upon several factors, including individual differences in the tendency towards resource conservation.

### Application of CET to a speech recognition task

In CET, the individual’s assessment of the importance of goal achievement, and the size of their resource pool, determines the magnitude of the ‘potential driving force’ (Kruglanski et al. 2012). Within a speech recognition task, we propose that the resource pool relates to two types of cognitive skills that have received particular scrutiny in the context of speech perception: working memory resources (Rönnberg 2003; Rönnberg et al. 2013; Rönnberg, Holmer, and Rudner 2019) and linguistic skills operationalised as lexical decision-making ability (Kaandorp et al. 2016).

In CET, actual effort expenditure is limited by a restraining force consisting of three additive components: (a) task demands, (b) alternative goals, which compete with the target activity for resources, and (c) resource conservation (Kruglanski et al. 2012). Resource conservation is high in individuals who have a high need for closure (Kruglanski and Webster 1996; Roets and van Hiel 2008; Viola et al. 2015), and based on previous studies (e.g. Beh 1990; Hinsz and Jundt 2005; Capa, Audiffren, and Ragot 2008), may be low in individuals who have a motivational style more focussed on achieving success (i.e. achievement motivation). Hence, considering these personality traits may help to account for differences in LE studies.

Another advantage of using CET over FUEL is that it makes quantifiable predictions about the likely effects of manipulating task demands and motivation on effort and performance in the context of driving and restraining forces. A strong driving force, for example, is expected to permit the use of more effective means to goal attainment and lead to better performance, though at a cost in terms of effort. A strong restraining force, on the other hand, restricts the use of these resource-heavy strategies and hence results in poorer performance. In the present study, our main aim was to investigate how listening demands and motivation (operationalised as financial reward) regulate listening effort in a speech recognition task, in the context of CET. Our secondary aims were to investigate the possible moderating influence of (i) resource conservation (operationalised as need for closure and individual differences in achievement motivation) and (ii) resource pool capacity (operationalised as working memory span and lexical-decision-making ability) on the relationship between motivation and listening demands. We chose different types of outcome measures to reflect the multidimensionality of LE and to test whether CET predictions apply to the subjective (self-reported ratings of LE and likelihood of ’giving up’, or avoidance, as described in Picou and Ricketts (2014)) and objective (correct response rate, RT) outcome measures used in the present study. We made the following predictions:

We predicted that listening demands would show a main effect, with higher demands resulting in lower correct response rates, longer RTs and higher subjective ratings of LE and likelihood of ‘giving up’. We also predicted a main effect of reward, with high value reward expected to motivate participants more than low, and result in higher correct response rates, longer RTs, higher self-rated LE and lower self-rated likelihood of giving up.
Interactions between listening demands and reward were predicted for the correct response rate, RTs and subjective ratings of LE and likelihood of giving up, driven by a greater motivational influence of financial reward under higher listening demands. In addition, we hypothesised that differences in the resource pool capacity (measured by working memory span and lexical-decision making ability) and in the resource conservation aspect of the restraining force (need for closure and individual differences in achievement motivation) would predict the correct response rate and subjective and behavioural measures of LE. We expected measures of resource conservation and measures of resource pool capacity to interact with listening demands and reward for behavioural and self-report LE outcomes.

Methods

Participants

To be eligible, participants needed to be between 18 and 35 years old, with normal hearing, normal or corrected-to-normal vision and no previous neurological issues or speech problems. Twenty-four (18 female) NH native-English speaking adults participated in the study, ranging from 19 to 33 years of age (median = 23). This sample size is sufficient to achieve 80% power (1 - b = .80, a = .05) for a medium effect size (f = .25) for a 2 x 2 repeated-measures factorial design (Faul et al. 2009) for each of the four main outcome measures (see Procedures and Data Analysis). Prior to taking part, participants were informed that the purpose of the study was to understand whether a person’s motivation to complete a listening task changes the amount of effort they use. Participants were compensated for their time with a £15 honorarium and were informed that they would have the chance to earn additional performance-based rewards by answering the questions correctly, in order to incentivise maximal effort exerted during the task (see Speech Recognition Task section below). The study was reviewed and approved by the University of Manchester Research Ethics Committee (approval number: 2019-6493-10583) and pre-registered with the Open Science Framework https://osf.io/6x7pd?view_only=d91bf9c111124fc2ab1fc6c52893182f.

Hearing screening

Each participant was screened using otoscopy, tympanometry and pure-tone audiometry to ensure they met the eligibility criteria for participation. All participants had bilateral NH (≤20 dB HL for test frequencies of 250, 500, 1000, 2000, 4000 and 8000 Hz) (British Society of Audiology (BSA) 2017) and reported no recent ear infections or surgery, previous neurological issues or speech problems.

Materials

Speech recognition task: Stimuli

A speech recognition task using degraded sentences was chosen as these types of tasks are effective in eliciting LE that can be measured using RTs (e.g. Gatehouse and Gordon 1990; Pals et al. 2015). Ninety Harvard IEEE sentences (Rothauser et al. 1969), spoken by a male speaker, were used as the speech materials. Speech intelligibility was modified using vocoding, an effective way to manipulate the intelligibility of speech in a controlled manner (Druilman, Festen, and Plomp 1994; Shannon et al. 1995). Vocoding has been shown to affect subjective and objective measures of LE (McMahon et al. 2016; Winn 2016).

Vocoded stimuli were created using a custom algorithm in Matlab (The Mathworks R 2018a). Speech stimuli were processed using a 2-band (high listening demands) or a 3-band (moderate listening demands) tone vocoder, with the frequency of each vocoder band logarithmically spaced between 80 and 8000 Hz. Two and 3 bands were chosen based on pilot testing that resulted in mean correct response rates of around 80% for the moderate listening demands condition and around 50% for the high listening demands condition, using the speech recognition task described below. The carrier frequencies were 225, 1047 and 4861 Hz for the 3-band vocoder and 440 Hz and 4440 Hz for the 2-band vocoder. The temporal envelope of the output of each channel was extracted using half-wave rectification and smoothing (using a low-pass filter with a cut-off frequency of 300 Hz) and used to modulate a sinusoidal carrier with a frequency equal to the centre frequency of the vocoder band. The signals within each band were then summed to produce tone-vocoded sentences.

Speech recognition task: Procedure

Figure 1 shows that in each trial of the speech recognition task, the sentence was presented twice. In the moderate listening demands condition, the first presentation (‘cue’) of the sentence was vocoded to produce a moderate degree of intelligibility (3-band vocoder) followed by a second presentation (‘target’) vocoded for low speech intelligibility (2-band vocoder). In the high listening demands condition, the sentence was always presented (both ‘cue’ and ‘target’) at a low intelligibility level (2-band vocoder). Thus, the second presentation (‘target’) of the

![Table](Image)

**Figure 1.** Depiction of a typical ‘high’ reward trial. In ‘low’ reward trials the pre-trial screen informed participants that the reward was £0.25 rather than £2.50.

![Diagram](Image)
speech sentence was always generated with a 2-band vocoder and was therefore identical in terms of its physical properties in both the high and moderate listening demands conditions, with only the ‘cue’ sentence changing in terms of its physical properties and perceived intelligibility. Our approach dissociates the perceptual effects of changes in speech intelligibility and LE from acoustical differences that can be used to vary listening demands. This is achieved by manipulating the perceived intelligibility of identical speech stimuli through prior exposure, that is, vocoded speech that is initially relatively unintelligible can become more intelligible after participants are exposed to an intelligible version of the same speech stimulus (e.g. Davis et al. 2005; Millman, Johnson, and Prendergast 2015). The use of identical speech stimuli that manipulate listening demands and LE could be particularly advantageous in interpreting changes in objective (physiological) measures of LE.

To future-proof the design for potential physiological testing, an assessment method for speech intelligibility was chosen that minimised movement-related noise caused by overt verbal responses. A test word from the sentence was selected randomly from either the beginning, middle or end of the sentence to ensure participants had to listen to the entire sentence. Of the 80 test words, 27 were selected from the beginning, 27 from the middle and 26 from the end of the sentence. Participants were asked to select, using a mouse, which word they had heard within the preceding sentence from amongst five foils presented as a 6-word visual grid (see Figure 1). The mouse cursor returned to the middle of the screen when the visual grid was displayed. The location of the test word varied randomly within the 6-word grid, with an equal chance of the test word appearing in any of the 6 positions. All five foils were either phonologically or semantically related to the test word or other foils. For instance, the sentence ‘The loss of a second ship was hard to take’ and the test word ‘take’ had the following foils: phonological foils related to ‘take’ (‘talk’, ‘tale’); a semantic foil for ‘take’ (‘accept’); phonological or semantic foils for other foils (‘except’, ‘tell’). An online rhyming dictionary, rhymezone.com, was used to select phonologically and semantically related foils (>90 similarity rating). The six options were presented immediately after the speech presentation to minimise memory requirements.

**Main outcome measures**

**Correct response rate and RTs**

Participants were asked to respond as accurately and quickly as possible. The percentage of correct responses (correct response rate) and the average speed of responses (RTs) were measured. Mean RTs were computed inclusive of incorrect trials to avoid data loss (Houben, van Doorn-Bierman, and Dreschler 2013). Mean RTs for incorrect trials were 3.1 s longer than RTs for correct trials ($t(748) = 16.48, p < .001$), but excluding incorrect trials from the analysis did not change the overall pattern of results.

**Subjective ratings of LE and likelihood of giving up**

After each trial in the speech recognition task (see Figure 1), the monitor displayed two consecutive questions to gauge subjective LE and the likelihood of giving up: ‘How hard did you work to understand what was said?’ and ‘How likely would you be to give up or just stop trying?’ We will refer to these measures as self-rated ‘work’ and ‘giving up’, respectively. The wording used to elicit these self-report ratings was almost identical to the wording used by Picou and Ricketts (2014, 2018) and is based upon questions from the Speech, Spatial and Qualities Hearing Scale (Gatehouse and Noble 2004). Participants provided subjective ratings, using a mouse, on a visual scale between 0 (‘not at all’) to 100 (‘very’).

**Other outcomes measures**

**NASA task load index**

In the NASA Task Load Index (NASA-TLX; Hart and Staveland 1988), participants are asked to rate how mentally, physically and temporally demanding they found a recently completed task. Additionally, participants were asked to give ratings on their perceived performance level, how much effort they used and how frustrating they found the task. Rating scales run between 1 (‘very low’) and 20 (‘very high’), except for self-rated performance for which the scale runs from 1 (‘failure’) to 20 (‘perfect’). Participants completed the NASA-TLX immediately after completing all trials. These ratings were collected to gain an overall picture of effort levels and perceptions of the task to aid interpretations of other analyses.

**Visual search task**

A target word was displayed visually and participants were instructed to select the target word in the 6-word grid as quickly as possible. The mean visual search RT was calculated based on 20 trials. As items within the 6-word response grid used for the speech recognition task were not equally spaced (i.e. selection of the outer items required a slightly greater mouse movement), the mean visual search RT for each participant was used to account for physical differences in the spacing of the items in the 6-word grid.

**Covariate measures**

**Motivational personality traits**

Achievement motivation was measured using the Personal Mastery and Competitive Excellence subscales of the Motivational Trait Questionnaire (Heggestad and Kanfer 2000). Both of these sub-scales index achievement-orientated traits: individuals scoring high in personal mastery strive to maximise their performance even for challenging tasks, whilst individuals high in competitive excellence strive to achieve a level of success above their peers. The personal mastery section has 16 items, and statements include: ‘I set goals as a way to improve my performance’. The competitive excellence section has 13 items, and statements include ‘Even in non-competitive situations, I find ways to compete with others’. Statements were rated between 1 (very untrue of me) and 6 (very true of me).

Need for closure was measured using the Need for Closure Scale (Roets and van Hiel 2011, updated from the original version written by Webster and Kruglanski 1994), which indexes a person’s closed-mindedness, dislike of uncertainty and preference for order and predictability (Roets et al. 2015). The scale has 15 items, example items include ‘I don’t like situations that are uncertain’. Participants rated these statements between 1 (strongly disagree) and 6 (strongly agree).
Cognitive tests
All cognitive tests were carried out using Inquisit 5 (Millisecond Software 2015). Working memory was assessed using an auditory version of the backwards digit span test. Participants were presented with a series of digits and asked to recall them in reverse order. Responses were recorded using a computer keyboard. Participants received two practice trials prior to the main assessment. Participants were initially presented with a 2-digit sequence. Subsequently, the sequence length was adjusted based on performance. Correct recall increased the length of the sequence by 1; failing to recall the sequence correctly after two attempts reduced the sequence length by 1. The backwards digit span was defined as the maximal sequence length of correctly recalled digits after 14 trials.

Linguistic ability was assessed using a lexical decision-making task. Participants were presented with 4 or 5 letter strings and had to indicate whether the strings made up words or non-words as quickly and accurately as possible. Participants recorded their responses via a computer keyboard to yield the lexical decision RT. The task consisted of a practice block containing 6 trials (3 non-words and 3 words in random order), followed by 52 test trials (consisting of 26 words and 26 non-words presented randomly). In each trial, a fixation cross was presented for 700 ms, followed by the stimulus for 250 ms and then a blank screen. The mean RT was calculated for correct trials only.

Procedures
All tasks were completed in a single testing session, which lasted around 1 hour. For the speech recognition task, participants were seated in a sound-attenuated booth facing a computer monitor and given task instructions. During each trial, vocoded sentences were presented diotically at a fixed level of 65 dB(A) via loudspeakers at ±45° azimuth.

After a practice block consisting of 10 trials, 80 test trials were presented in 8 blocks of 10 trials each. There were four high-reward (£2.50) and four low-reward (£0.25) blocks presented in random order. No explanation as to why some trials were worth more than others was provided. Prior to each block, participants were informed that they would receive a financial bonus for answering 6 or more items correctly over the next block of 10 trials. Each block consisted of five trials with moderate listening demands and five trials with high listening demands, presented in random order. Feedback on the performance and the associated award was not given until the end of the experiment, to dissociate the effects of financial reward from mood-related changes in effort, which may occur when participants are given trial-by-trial feedback (Carver 2006; Koelewijn et al. 2018).

After the speech-recognition task, participants completed the visual search task. Following this, participants were asked to complete the NASA-TLX (Hart and Staveland 1988) and the personality questionnaires. Finally, participants were asked to complete the backwards digit span and the lexical decision-making tasks.

Data analyses
Prior to statistical analysis, correct response rates, self-rated ‘work’ and self-rated ‘giving up’ were converted to rationalised arcsine units (RAU) (Studebaker 1985). To remove outliers from the RT data, RTs further than three standard deviations from the mean for each participant were removed (Picou, Charles, and Ricketts 2017). A log10 transformation was then applied to the RTs to meet the assumption of normality for parametric statistics. For each dependent variable (correct response rate, RT, self-rated ‘work’, self-rated ‘giving up’), a repeated-measures analysis of variance (ANOVA) with two within-subject factors, listening demands (moderate/high) and financial reward (low/high) was conducted.

Linear mixed modelling was carried out to investigate whether cognitive skills and personality traits predicted outcomes from the speech recognition task. Statistical analyses were run in R version 3.5.1 (R Core Team 2018), using RStudio 1.1.453 and the nlme package (Pinheiro et al. 2020). For each outcome measure from the speech recognition task (correct response rate, RTs, self-rated “work” and self-rated “giving up”) exploratory mixed models were fitted. Eight fixed effect predictors were included: listening demands, financial reward, an interaction term for listening demands and financial reward, backwards digit span, mean lexical decision RT, mean need for closure score, and total scores on the Personal Mastery and Competitive Excellence subscales of the Motivational Trait Questionnaire. Participants were included as a random effect in all mixed models. We used a backwards stepwise procedure (Pinheiro and Bates 2000) to prune the initial model in such a way that higher-level interaction terms only remained if they improved the model fit.1

For each significant cognitive and personality main effect in the exploratory models, we conducted further analyses to investigate whether these predictors interacted with financial reward or listening demands. The full model included main effects for listening demands, reward and the cognitive or personality effect under investigation, plus all first- and second-level interaction effects and was subsequently pruned in the manner described above.

Results
Speech recognition and LE measures
Figure 2 shows the results of the speech recognition task and the associated measures of LE. The correct response rates (% correct) for the speech recognition task are shown in Figure 2(a). A repeated-measures ANOVA with two factors (moderate/high listening demands and low/high financial reward), showed a significant effect of listening demands \( F(1,23) = 53.76, \quad MSE = 146.49, \quad p < .001, \quad \eta_p^2 = .70 \) on the correct response rate, with a higher mean correct response rate in the moderate compared with the high listening demands condition collapsed across reward condition (moderate: mean = 69.6%, SEM = .020; high: mean = 50.2%, SEM = .031). There was no significant effect of financial reward on the correct response rate \( F(1,23) = .296, \quad MSE = 115.49, \quad p = .592, \quad \eta_p^2 = .013 \) and no significant interaction between listening demands and financial reward \( F(1,23) = .015, \quad MSE = 85.55, \quad p = .902, \quad \eta_p^2 = .001 \).

Figure 2(b) shows mean RTs (log10(s)) for the speech recognition task. A repeated-measures ANOVA conducted on the RTs showed a significant effect of listening demands \( F(1,23) = 18.02, \quad MSE = .01, \quad p < .001, \quad \eta_p^2 = .44 \) with a slower mean RT in the high listening demands condition, compared with the moderate listening demands condition collapsed across reward condition (high: mean = .788 log10(s), SEM = .023; moderate: mean = .714 log10(s), SEM = .024). There was no significant effect of financial reward on RTs \( F(1,23) = 1.83, \quad MSE = .01, \quad p = .190, \quad \eta_p^2 = .074 \) and the interaction between listening demands and financial reward was non-significant \( F(1,23) = .250, \quad MSE = .01, \quad p = .622, \quad \eta_p^2 = .011 \). To control for physical
differences in the spacing of items in the 6-word grid, we calculated a mean visual search RT for each participant (see Visual Search Task in Methods) and subtracted this value from the mean RTs gathered in the listening task in each condition. Running the ANOVA on the adjusted listening task RTs revealed the same pattern of results (i.e. a significant effect of listening demands ($F(1,23) = 19.17, MSE = 22.99, p < .001, \eta^2_p = .455$), a non-significant effect of motivation ($F(1,23) = 1.77, MSE = 2.64, p = .197, \eta^2_p = .071$) and a non-significant interaction ($F(1,23) = .29, MSE = .27, p = .594, \eta^2_p = .013$).

Figs. 2C and 2D show ‘work’ and ‘giving up’ self-report ratings (0–100%) for the speech recognition task, respectively. For self-rated ‘work’, a repeated-measures ANOVA showed a significant effect of listening demands ($F(1,23) = 21.74, MSE = 57.67, p < .001, \eta^2_p = .49$) with a higher mean ‘work’ rating when listening demands were high compared with moderate, collapsed across reward condition (high: mean = 65.59%, SEM = 3.858; moderate: mean = 58.50%, SEM = 4.076). There was a significant main effect of financial reward ($F(1,23) = 6.94, MSE = 23.85, p = .015, \eta^2_p = .23$) with a higher mean “work” rating for high reward compared to low reward collapsed across listening demand conditions (high: mean = 63.43%, SEM = 3.829; low: mean = 60.67%, SEM = 3.829). No significant interaction between reward and listening demands was measured for ‘work’ ratings ($F(1,23) = .258, MSE = 13.32, p = .616, \eta^2_p = .011$). A significant main effect of listening demands was measured for self-rated ‘giving up’ ($F(1,23) = 24.51, MSE = 25.87, p < .001, \eta^2_p = .52$), driven by a higher mean ‘giving up’ rating for high compared to moderate listening demands collapsed across reward conditions (high: mean = 31.94%, SEM = 5.001; moderate: mean = 27.73%, SEM = 4.644). Financial reward did not have a significant effect on self-rated ‘giving up’ ($F(1,23) = .942, p = .342, \eta^2_p = .039$) and the interaction between financial reward and listening demands was not significant ($F(1,23) = .710, MSE = 10.44, p = .408, \eta^2_p = .030$).

**NASA-TLX**

Figure 3 shows mean ratings for each subscale of the NASA-TLX (Hart and Staveland 1988). Participants reported high levels of effort, mental demand and frustration after completing the speech recognition task, suggesting that participants engaged with the task and expended LE.

**Cognitive and personality measures**

Group means, standard deviations and ranges for motivational traits and cognitive abilities are shown in Table 1. Based on means/ranges in previous studies (e.g. Viola et al. 2015), all participants would be classified as low in need for closure. Means and standard deviations for the Motivational Trait Questionnaire subscales were similar to Hinsz and Jundt (2005). Means and standard deviations for backwards digit span were very similar to those recorded for young NH participants in Woods et al. (2011). Means and standard deviations for lexical decision RTs were similar to the NH young participants in Strand et al. (2018). Pearson’s correlation coefficients between motivational traits (need for closure, competitive excellence and personal mastery) and cognitive skills (backwards digit span and lexical decision RT) were non-significant and small (all $r < .2$ except competitive excellence and need for closure where $r = .44$, data not shown).

**Multi-level modelling of cognitive and personality factors**

Table 2 shows exploratory mixed models for each outcome measure, which included listening demands, reward, demand*reward interaction, backwards digit span, lexical decision RT and totals for the Competitive Excellence and Personal Mastery...
Table 1. Group means, standard deviations and ranges for personality traits/cognitive abilities.

| Trait                      | Mean (S.D.) | Range      |
|----------------------------|-------------|------------|
| Competitive excellence     | 65.61 (10.50)| 26 – 65   |
| Personal mastery          | 74.25 (11.33)| 51 – 96   |
| Need for closure           | 3.80 (.46)  | 2.87 – 4.80|
| Backwards digit span       | 6.42 (97)   | 4 – 8     |
| Lexical decision RT (ms)  | 694.51 (176.73)| 512.60 – 1121.26|

subsections of the Motivational Trait Questionnaire and the Need for Closure Scale as fixed effects. For correct response rate, alongside a significant effect of listening demands ($F(1, 1893) = 81.23, p < .001, \eta^2_g = .04$) we found a significant fixed effect of backwards digit span ($F(1, 18) = 7.87, p = .01, \eta^2_p = .01$). Our analysis showed that the best fitting model for the correct response rate consisted of listening demands, reward and backwards digit span, of which listening demands ($F(1, 1894) = 81.27, p < .001, \eta^2_p = .02$) and backwards digit span ($F(1,22) = 6.62, p = .02, \eta^2_p = .01$) were individually significant fixed effects. Hence, backwards digit span did not interact with either listening demand or financial reward to affect correct response rate.

For RTs, in addition to a significant fixed effect of listening demands ($F(1, 1870) = 41.29, p < .001, \eta^2_g = .02$), we found significant fixed effects for competitive excellence ($F(1, 18) = 5.74, p = .03, \eta^2_p = .02$) and lexical decision RT ($F(1,18) = 6.53, p = .02, \eta^2_p = .02$). When exploring potential interaction effects of competitive excellence on RTs, the best-fitting model showed individually significant fixed effects for listening demands ($F(1, 1871) = 41.28, p < .001, \eta^2_p = .02$) and competitive excellence ($F(1,22) = 4.27, p = .05, \eta^2_p = .02$) only. When exploring potential interaction effects of lexical decision RTs on RTs, the best-fitting model showed that only listening demands ($F(1, 1871) = 41.29, p < .001, \eta^2_p = .02$) and lexical decision RT ($F(1,22) = 5.23, p = .03, \eta^2_p = .02$) were individually significant fixed effects.

Only fixed effects of listening demands and reward were found for self-rated work (demands: $F(1, 1893) = 65.61, p < .001, \eta^2_p = .02$; reward: $F(1, 1893) = 11.30, p < .001, \eta^2_p = .03$) and giving up (demands: $F(1, 1893) = 56.51, p < .001, \eta^2_p = .01$; reward: $F(1,1893) = 7.40, p = .01, \eta^2_p = .01$). Therefore no further interactions were explored.

Discussion

This study evaluated the relationship between listening demands and motivation in a speech recognition task in the context of a multi-factorial model from the wider field of effort research, i.e., CET (Kruglanski et al. 2012). We manipulated motivation by varying financial reward and listening demands by varying the degree of degradation of the vocoded speech presented to listeners. We measured the effects of these manipulations on four main outcomes (correct response rate, RTs, self-rated work, self-rated giving up). We also considered the modulating effects of personality factors and cognitive skills. The manipulations and co-varying factors, as well as the resulting hypotheses, reflect predictions made by CET.

The prediction of a main effect of listening demands was supported. Varying prior knowledge of tone-vocoded speech was found to be an effective way of manipulating listening demands: high listening demands led to significantly decreased correct response rates, increased RTs and increased self-rated “work” and “giving up”. These findings are consistent with CET (Kruglanski et al. 2012) and are also in line with FUEL (Pichora-Fuller et al. 2016).

We also found the predicted main effect of reward, consistent with CET, which stipulates that financial reward increases motivation, resulting in a stronger driving force and greater mobilisation of effort to counteract the restraining force. It is important to note, however, that according to the mixed model (Table 2), the effects of financial reward were limited to self-rated work and giving up and did not extend to increased correct response rates or changes in RTs. It is possible these results did not reflect LE but instead were due to demand characteristics of the experiment, that is, participants realising that greater effort was expected in high reward trials.

Although we cannot rule out demand effects, we propose two alternate interpretations for why financial reward affected only self-rated but not behavioural outcomes. First, LE may be a multi-dimensional concept (McMahon et al. 2016; Strauss and Francis 2017; Hughes et al. 2018; Strand et al. 2018; Alhanbali et al. 2019) where some measures, for example, self-report, show an effect of LE and others, for example, behavioural outcomes, do not. In a similar vein, other studies have shown physiological effects in response to increased LE under conditions of high financial reward but no significant behavioural effects (e.g. Richter 2016; Koelwijn et al. 2018). Self-rated LE measures may also be more sensitive to the effects of motivation than behavioural measures (Pichora-Fuller et al. 2016; Herrmann and Johnsrude 2020), which might explain why RTs did not appear to be sensitive to the financial reward manipulation used in this study. Second, RTs are sensitive to how motivation is operationalised: Weis et al. (2013) found differential effects on RTs in an auditory discrimination task depending upon whether the financial motivator was presented to participants as a reward (starting from zero and gaining money for correct answers) or a punishment (starting from maximum and losing money for incorrect responses).

The interaction between listening demands and financial reward was non-significant cf. CET, FUEL and MIT. Richter, Gendolla, and Wright (2016) suggest a number of extensions to MIT (Breym and Self 1989) that may limit the greater influence of motivation at higher demands, some of which may explain the lack of interactive effects in this study. These extensions include fatigue level, mood, and participant perceptions of their ability to succeed at a task. In the present study, despite scoring well above chance (~50% correct) in the high listening demands condition, some participants may have perceived the task as too difficult and hence offering greater reward would have little impact upon effort investment. This may also account for the very high levels of frustration recorded on the NASA-TLX (Figure 3). Coupled with differences in task design, this may also explain why our results conflict with Zhang et al. (2019) who found an interaction between demands and reward on performance in NH listeners at high correct response rates (~70-85%).

We predicted that differences in the resource pool capacity (measured by backwards digit span and lexical decision RTs) and in the resource conservation aspect of the restraining force (i.e. need for closure and inter-individual differences in achievement motivation) would impact upon the correct response rate and subjective and behavioural measures of LE in our speech recognition task. A significant main effect of backwards digit span was found for the correct response rate. This result suggests that participants with greater resources performed better, consistent with the driving force component of CET. This finding also supports the ELU model (Rönnerberg 2003; Rönnerberg et al. 2013;
Rönning, Holmer, and Rudner 2019), that is, working memory resources are recruited during effortful listening to resolve mismatches between input and representations stored in long-term memory. However, note that individual differences in backwards digit span did not predict behavioural (RT) or self-rated LE, a result that we discuss further in the limitations section (see below).

In contrast, lexical decision RT predicted only RTs from the speech recognition task and not correct response rate. Moreover, the direction of this association did not follow CET predictions. Specifically, CET predicts greater resources (here we assumed speech recognition task and not correct response rate). Moreover, the result that we discuss further in the limitations section (see below).

| Task          | Demands | Reward | Demands* | Reward | Backwards | Lexical decision | Competitive | Personal | Need for Closure |
|---------------|---------|--------|----------|--------|-----------|------------------|-------------|----------|-----------------|
| Correct rate  | 81.23** | 0.34   | 0.04     | 7.87*  | 3.77      | 1.46             | 1.18        | 1.73     |
| RT            | 41.29** | 2.55   | 1.29     | <0.01  | 6.53*     | 5.79             | 0.69        | 4.16     |
| Work          | 65.61** | 11.30**| 0.15     | 1.91   | 0.59      | <0.01            | 0.09        | 1.61     |
| Give up       | 56.51** | 7.40*  | 0.53     | 1.54   | 1.66      | 1.17             | 0.09        | 1.32     |

Bold and **p < .001; *p < .05.

Limitations of the present study

We did not find a consistent effect of financial reward on the main outcome measures in the present study, whereas research which informed CET predictions includes experiments where motivation in a listening task was manipulated using financial reward (e.g. Bijleveld, Custers, and Aarts 2009). However, the effectiveness of extrinsic rewards when manipulating motivation has been questioned. Previous meta-analyses have concluded that offering tangible rewards undermines a person’s intrinsic motivation, that is, their desire to engage with interesting tasks to the best of their ability (Rummel and Feinberg 1988; Wiersma 1992; Tang and Hall 1995; Deci, Koestner, and Ryan 1999). A performance-contingent reward may erode a person’s perceived autonomy and competence at a task if they attribute their performance to be due to the reward rather than their own interest (Lepper, Greene, and Nisbett 1973; Deci and Ryan 1985). It is, therefore, possible that financial reward may have demotivated some participants in the present study.

More granular aspects of study design may also impact upon the effectiveness of the motivating variable. We presented a high (£2.50) versus low (£0.25) reward for achieving a correct response rate of ≥60% per every 10 trials. Other studies used a much higher reward threshold e.g. 90% (Richter 2016) which may have encouraged greater effort, although setting the threshold to gain a reward too high may discourage effort investment if the goal is perceived as impossible (Brehm, Greene, and Nisbett 1973; Deci and Ryan 1985). In addition, participants could feasibly exert high effort in every trial, regardless of reward condition, to maximise the overall amount of reward they received. Introducing the need to strategically allocate resources (e.g. Zhang et al. 2019), may result in a more effective manipulation of motivation.

The lack of significant interactions between listening demands and financial reward, and the null effects for the resource pool and resource conservation covariates on our main outcome measures, may be explained by a lack of statistical power and/or a lack of variability in our resource pool measures. Our sample size reflects the number of participants needed to identify main effects on the main outcome measures. The current results will inform appropriate sample sizes in future studies to investigate the interactions predicted by CET/MIT. Such future research can then clarify whether the current null effects were due to methodological limitations of the present study (e.g. lack of extreme groups for personality traits, the possibility that participants perceived the listening task to be too hard), or whether CET is not appropriate for predictions in this particular context.
Conclusions
The present study shows that manipulating prior knowledge by using vocoded speech is a feasible way of varying speech intelligibility and associated measures of LE (RTs, self-ratings) in young, NH listeners. The effects of offering financial reward on LE were more complex: changes in subjective ratings of ‘work’ and ‘giving up’ with increased financial reward were not mirrored by increased correct response rates or greater LE investment, as measured by RTs. We found only partial support for CET predictions which are intended to apply to ‘all instances of goal-directed thinking’ (p. 3, Kruglanski et al. 2012). It is unclear whether this is due to the limitations of financial reward as a manipulator of motivation or the multi-dimensional nature of LE (McMahon et al. 2016; Pichora-Fuller et al. 2016; Strauss and Francis 2017; Hughes et al. 2018; Strand et al. 2018; Alhanbali et al. 2019). The results of the exploratory analyses presented here suggests the influence of personality and cognitive skills on effortful listening and their interaction with listening demands and motivation is small.

Note
1. The Akaike Information Criterion (AIC) was used to estimate the fit of each model. Comparisons were made between the AIC of a model containing a particular interaction effect and a model excluding this interaction term while keeping all other terms identical. The pruned model with the lowest AIC value was then compared with the unpruned model for this level. If the AIC of the pruned model was lower, the pruned model was carried forward to the next stage of analysis and set as the new base model for pruning. If the AIC of the pruned model was higher than the unpruned model, indicating a worse fit, an ANOVA was conducted to compare both model fits. If the pruned model was not significantly worse, it was carried forward as the new base model for pruning. Following this procedure, interaction terms were progressively eliminated, until only one remained. The final model was established by using ANOVA to compare this model to a model consisting of only the three main effects. ML (maximum likelihood) estimation was used for the stepwise comparisons and upon establishing the final model, fixed effects were calculated using REML (restricted maximum likelihood) estimation i.e. a modelling approach similar to Heinrich, Ferguson, and Mattys (2019) and Knight and Heinrich (2017, Knight and Heinrich 2018).

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This work was funded (PJC) and supported (AH, KJM, REM) by the NIHR Manchester Biomedical Research Centre [BRC-1215-20007].

ORCID
Peter J. Carolan http://orcid.org/0000-0003-4386-0310
Antje Heinrich http://orcid.org/0000-0002-5940-622X
Kevin J. Munro http://orcid.org/0000-0001-6543-9098
Rebecca E. Millman http://orcid.org/0000-0001-8606-0167

Data availability statement
The data that support the findings of this study are available from the corresponding author, PJC, upon request.

References
Alhanbali, S., P. Dawes, R. E. Millman, and K. J. Munro. 2019. “Measures of Listening Effort Are Multidimensional.” Ear and Hearing 40 (5): 1084–1097. doi:10.1097/AUD.0000000000000697.
Beh, H. C. 1990. “Achievement-Motivation, Performance and Cardiovascular Activity.” International Journal of Psychophysiology: official Journal of the International Organization of Psychophysiology 10 (1): 39–45. doi:10.1016/0167-8760(90)90043-D.
Bilevicius, E., R. Custers, and H. Aarts. 2009. “The Unconscious Eye Opener: Pupil Dilation Reveals Strategic Recruitment of Resources upon Presentation of Subliminal Reward Cues.” Psychological Science 20 (11): 1313–1315. doi:10.1111/j.1467-9280.2009.02443.x.
Brehm, J. W., and E. A. Self. 1989. “The Intensity of Motivation.” Annual Review of Psychology 40: 109–131. doi:10.1146/annurev.ps.40.020189.000545.
British Society of Audiology (BSA) 2017. Recommended Procedure: Pure-Tone Air-Conduction and Bone-Condition Audiometry with and without Masking. British Society of Audiology. Retrieved from: https://www.bsa.org
Capa, R. L., M. Audiffren, and S. Ragot. 2008. “The Effects of Achievement Motivation, Task Difficulty, and Goal Difficulty on Physiological, Behavioral, and Subjective Effort.” Psychophysiology 45 (5): 859–868. doi:10.1111/j.1469-8986.2008.00675.x.
Carrillo de la Peña, M. T., and F. Cadeveira. 2000. “The Effect of Motivational Instructions on P300 Amplitude.” Clinical Neurophysiology 30 (4): 232–239. doi:10.1016/S0987-7053(00)00220-3.
Carver, C. S. 2006. “Approach, Avoidance, and the Self-Regulation of Affect and Action.” Motivation and Emotion 30 (2): 105–110. doi:10.1007/s11031-006-9044-7.
Cohen, J. 1988. Statistical Power Analysis for the Behavioural Sciences. New York: Academic Press. doi:10.1016/C2013-0-10517-X.
Davis, M. H., I. S. Johnsrude, A. Hervais-Adelman, K. Taylor, and C. McGettigan. 2005. “Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence from the Comprehension of Noise-Vocoded Sentences.” Journal of Experimental Psychology. General 134 (2): 222–241. doi:10.1037/0022-358X.134.2.222.
Deci, E. L., R. Koestner, and R. M. Ryan. 1999. “A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation.” Psychological Bulletin 125 (6): 627–668. doi:10.1037/0033-2909.125.6.627.
Deci, E. L., and R. M. Ryan. 1985. Intrinsic Motivation and Self-Determination in Human Behaviour. New York: Plenum.
Drullman, R., J. M. Festen, and R. Plomp. 1994. “Effect of Reducing Slow Temporal Modulations on Speech Reception.” The Journal of the Acoustical Society of America 95 (5 Pt 1): 2670–2680. doi:10.1121/1.40836.
Faul, F., E. Erdfelder, A. Buchner, and A. G. Lang. 2009. “Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses.” Behavior Research Methods 41 (4): 1149–1160. doi:10.3758/BRM.41.4.1149.
Gatehouse, S., and W. Noble. 2004. “The Speech, Spatial and Qualities of Hearing Scale (SSQ).” International Journal of Audiology 43 (2): 85–99. doi:10.1080/14992020400050014.
Hackley, S. A., and F. Valle-Inclán. 1998. “Automatic Alerting Does Not Speed Late Motoric Processes in a Reaction-Time Task.” Nature 391 (6669): 786–788. doi:10.1038/35849.
Hart, S., and L. Staveland. 1988. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research.” Advances in Psychology 52: 139–183. doi:10.1016/0166-4115(88)90011-3.
Heggestad, E. D., and R. Kanfer. 2000. “Individual Differences in Trait Motivation: Development of the Motivational Trait Questionnaire.” International Journal of Educational Research 33 (7-8): 751–776. doi:10.1016/S0883-0355(00)00049-5.
Henderson, M. T., and F. Cadeveira. 2000. “The Effect of Reducing Slow Temporal Modulations on Speech Reception.” The Journal of the Acoustical Society of America 95 (5 Pt 1): 2670–2680. doi:10.1121/1.40836.
Heinrich, A., M. A. Ferguson, and S. L. Mattys. 2019. “Effects of Cognitive Load on Pure-Tone Audiometry Thresholds in Younger and Older Adults.” Ear and Hearing 41 (4): 907–917. doi:10.1097/AUD.0000000000000812.
Herrmann, B., and L. Johnsrude. 2020. “A Model of Listening Engagement (MoLE).” Hearing Research 397: 108016. doi:10.1016/j.heares.2020.108016.
Hinzs, V. B., and D. K. Hundt. 2005. “Exploring Individual Differences in a Game-Setting Situation Using the Motivational Trait Questionnaire.” Journal of Applied Social Psychology 35 (3): 551–571. doi:10.1111/j.1559-1816.2005.tb02135.x.
Rothauser, E. H., W. D. Chapman, N. Guttman, M. H. L. Hecker, K. S. Nordby, H. Silbiger, G. E. Urbanek, and M. Weinstock. 1969. “IEEE Recommended Practice for Speech Quality Measurements.” IEEE Transactions on Audio and Electroacoustics 17 (3): 225–246.

Rummel, A., and R. Feinberg. 1988. “Cognitive Evaluation Theory: A Meta-Analytic Review of the Literature.” Social Behavior and Personality: An International Journal 16 (2): 147–164. doi:10.2224/sbp.1988.16.2.147.

Sankaran, S., E. Szumowska, and M. Kossowska. 2017. “When the going gets tough, the tough get going”: Motivation towards closure and effort investment in the performance of cognitive tasks.” Motiv Emot 41 (3): 308–332. doi:10.1007/s11031-017-9613-y.

Shannon, R., F. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. 1995. “Speech Recognition with Primarily Temporal Cues.” Science (New York, N.Y.) 270 (5234): 303–304. doi:10.1126/science.270.5234.303.

Strand, J. F., V. A. Brown, M. B. Merchant, H. E. Brown, and J. Smith. 2018. “Measuring Listening Effort: Convergent Validity, Sensitivity, and Links with Cognitive and Personality Measures.” Journal of Speech, Language, and Hearing Research: JSLHR 61 (6): 1463–1486. doi:10.1044/2018_JSLHR-H-17-0257.

Strauss, D. J., and A. L. Francis. 2017. “Toward a Taxonomic Model of Attention in Effortful Listening.” Cognitive, Affective and Behavioral Neuroscience 17 (4): 809–825. doi:10.3758/s13415-017-0513-0.

Studebaker, G. A. 1985. “A "rationalized" arcsine transform.” Journal of Speech and Hearing Research 28 (3): 455–462. doi:10.1044/jslhr.2803.455.

Tang, S. H., and V. C. Hall. 1995. “The over-Justification Effect: A Meta-Analysis.” Applied Cognitive Psychology 9 (5): 365–404. doi:10.1002/acp.2350090502.

The Mathworks R 2018a. Matlab. Natick, Massachusetts.

Viola, V., A. Tosoni, A. Brizi, I. Salvato, A. W. Kruglanski, G. Galati, and L. Mannetti. 2015. “Need for Cognitive Closure Modulates How Perceptual Decisions Are Affected by Task Difficulty and Outcome Relevance.” PLoS One 10 (12): e0146002. doi:10.1371/journal.pone.0146002.

Webster, D. M., and A. W. Kruglanski. 1994. “Individual Differences in Need for Cognitive Closure.” Journal of Personality and Social Psychology 67 (6): 1049–1062. doi:10.1037/0022-3514.67.6.1049.

Weis, T., S. Puschmann, A. Brechmann, and C. M. Thiel. 2013. “Positive and Negative Reinforcement Activate Human Auditory Cortex.” Frontiers in Human Neuroscience 7: 842. doi:10.3389/fnhum.2013.00842.

Wiersma, U. J. 1992. “The Effects of Extrinsic Rewards in Intrinsic Motivation: A Meta-Analysis.” Journal of Occupational and Organizational Psychology 65 (2): 101–114. doi:10.1111/j.2044-8325.1992.tb00488.x.

Winn, M. B. 2016. “Rapid Release from Listening Effort Resulting from Semantic Context, and Effects of Spectral Degradation and Cochlear Implants.” Trends in Hearing 20: 233121651666972. doi:10.1177/2331216516669723.

Woods, D. L., M. M. Kishiyama, E. W. Lund, T. J. Herron, B. Edwards, O. Poliva, R. F. Hink, and B. Reed. 2011. “Improving Digit Span Assessment of short-term verbal memory.” J Clin Exp Neuropsychol 33 (1): 101–111. doi:10.1080/13803395.2010.493149.

Zhang, M., G. J. Siegle, M. R. McNeil, S. R. Pratt, and C. Palmer. 2019. “The Role of Reward and Task Demand in Value-Based Strategic Allocation of Auditory Comprehension Effort.” Hearing Research 381: 107775 doi:10.1016/j.heares.2019.107775.