

RESEARCH ARTICLE 

# Is L2 pronunciation affected by increased task complexity in pronunciation-unfocused speaking tasks?

Ingrid Mora-Plaza , Joan C. Mora , Mireia Ortega  and Cristina Aliaga-Garcia 

Department of Modern Languages and Literatures and English Studies, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, Barcelona, 08007, Spain

**Corresponding author:** Ingrid Mora-Plaza; Email: [imoraplaza@ub.edu](mailto:imoraplaza@ub.edu)

(Received 19 September 2023; Revised 15 June 2024; Accepted 03 July 2024)

## Abstract

This study examines the effects of task complexity on second language (L2) pronunciation accuracy and global pronunciation measures in pronunciation-unfocused tasks and assesses the relationship between acoustic and listener-based pronunciation measures. Eighty-two Catalan/Spanish learners of English performed simple and complex versions of a problem-solving monologic speaking task, for which the oral stops /p, t, k/ and vowel contrasts /i:/-/ɪ/ and /æ/-/ʌ/ were embedded in the lexical items used to perform the task. Pronunciation accuracy was gauged through acoustic measurements of laryngeal timing (voice onset time), vowel contrastiveness and nativelikeness (Mahalanobis distances), and native speakers' ratings of comprehensibility and accentedness. Results revealed detrimental effects of increased task complexity on the productions of oral stops and speech comprehensibility and accentedness; however, no consistent task complexity effects were found on vowel accuracy. The analysis also revealed an association between segmental accuracy and global dimensions of L2 speech.

**Keywords:** accentedness; comprehensibility; pronunciation; segmentals; task complexity

## Introduction

Previous research investigating second language (L2) learners' oral production within task-based language teaching (TBLT) has provided empirical support for the tenets of Robinson's (2001a, 2011) cognition hypothesis, namely that increasing cognitive task demands along resource-directing variables (e.g., causal reasoning) enhances learners' attention to linguistic form and may therefore result in more complex language production and increased lexical and grammatical accuracy, often at the expense of speaking fluency (e.g., Gilabert, Barón & Llanes, 2007; Ishikawa, 2008). Such complexity manipulations are presumed to have a positive impact on L2 learners' linguistic

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

performance. An alternative hypothesis, namely Skehan's limited attentional capacity model (LACM) (2009, 2015), suggests that the complexity and accuracy dimensions of oral production might compete for learners' limited attentional resources, and may not both be attended to, resulting in complexity–accuracy trade-offs (e.g., Michel, Révész, Shi & Li, 2019; Sample & Michel, 2015). However, none of these hypotheses have been sufficiently investigated in relation to L2 pronunciation. Pronunciation is, in fact, underrepresented in TBLT research (Gurzynski-Weiss, Long & Solon, 2017), which has primarily focused on the conceptualization and formulation stages of speech production, neglecting the phonological and phonetic aspects of pronunciation, which can be influenced by task complexity.

Positive effects of increased task complexity have been found for speech comprehensibility but not for accentedness (e.g., Crowther, Trofimovich, Saito & Isaacs, 2018; Gordon, 2021), and for L2 pronunciation accuracy for a subset of L2 vowels (Mora-Plaza, 2023; Solon, Long & Gurzynski-Weiss, 2017). However, strong empirical evidence of the benefits of manipulating task complexity for enhancing attention to phonetic form is still lacking, whereas it is well-attested for lexical (Gilbert *et al.*, 2009), grammatical (Révész, 2009) and pragmatic (Márquez & Barón, 2021) form. According to Kormos' (1999, 2000) attention and monitoring model of speech processing, conceptualizing the message during online tasks may necessitate particular attention, leaving few attentional resources for lexical, semantic, and phonological encoding. In the case of L2 communicative tasks, learners may be forced to focus on lexical and grammatical aspects during speech production, making it difficult for them to pay attention to pronunciation due to the limited attentional resources available to them during self-monitoring (Kormos, 1999). In fact, lexical and grammatical self-repairs have been found to outnumber phonological repairs in purely meaning-oriented tasks (Kormos, 2000). The fact that most repairs are lexical in meaning-oriented tasks is because they carry most of the relevant information in the message and making errors may result in serious misunderstandings.

The present study aims to extend this line of research by investigating task complexity effects on L2 pronunciation accuracy in pronunciation-unfocused tasks. Pronunciation is an important component of language competence affecting listeners' comprehensibility (i.e., ease of understanding) and facilitating effective communication. In addition, investigating task effects on L2 pronunciation accuracy will provide insights into the role of speaking tasks and task design in fostering learners' pronunciation skills. Oral productions elicited from first language (L1)–Spanish advanced learners of English performing simple and complex versions of a problem-solving monologic speaking task were analyzed acoustically to obtain voice onset time (VOT) measures of laryngeal timing accuracy for L2 voiceless oral stops (/p/, /t/, /k/), and contrastiveness and nativelikeness for difficult L2 vowels (/i:/, /ɪ/, /æ/, /ʌ/). In addition, native English native listeners' (NL) judgments of comprehensibility and accentedness were obtained as global measures of L2 pronunciation accuracy.

### **TBLT: task design and manipulation**

TBLT is an analytic approach to language acquisition in which learners are presented with holistic samples of language, which they are expected to analyze and infer the underlying rules by themselves. In such a process, directing learners' attention toward accuracy while maintaining the communicative value of tasks is central to language development (Long, 2015). The use of a wide variety of pedagogical procedures to draw

learners' attention to linguistic form (see Sudharshana, 2021, for a review) would enhance learners' ability to refine and restructure their interlanguage.

In TBLT, tasks are conceived as real-world communicative activities requiring learners' use of language (Van den Branden, 2006), hence, a meaning-driven work plan that learners have to accomplish by relying on their own linguistic and non-linguistic resources (Ellis, 2009). Tasks can be categorized as unfocused, aiming to offer learners opportunities for general communicative language use, or focused, intending to provide opportunities for communication using specific linguistic features (Ellis, 2009). Focused tasks have often been found to effectively direct learners' attention to the use of the target linguistic features under focus. Manipulations of task design variables include task types (e.g., narrative/instruction-giving/decision-making; Gilbert et al., 2009) interlocutor proficiency (e.g., low/high; Kim & McDonough, 2008), task mode (e.g., online/face-to-face; Baralt, Gurzynski-Weiss & Kim, 2016) and task complexity (e.g., simple/complex; Révész, 2009). Empirical research has found that such task manipulations can influence the linguistic complexity, accuracy, and/or fluency (CAF) of learners' oral performance and the development of L2 linguistic accuracy. In particular, TBLT is considered an effective methodology for developing lexico-grammatical (Baralt et al., 2016) and pragmatic linguistic targets (Márquez & Barón, 2021). Nevertheless, very little attention has been paid to how unfocused communicative tasks might affect L2 pronunciation and to what extent task design and manipulation (i.e., task complexity) can effectively direct learners' attention to linguistic targets beyond grammar, lexis, and pragmatics (Gurzynski-Weiss et al., 2017), such as pronunciation.

### **Task complexity and CAF in oral performance**

With the aim of grading and sequencing tasks in a principled way in a task-based syllabus, L2 researchers proposed a set of criteria for evaluating the complexity of a task supported by theoretical frameworks, and conducted empirical studies to investigate whether task complexity on L2 production was predicted by those theories. First, Skehan's (2009, 2015) LACM, founded on theories of working memory and speech production, conceptualizes attention as a single volume that runs out of resources (Kahneman, 1973). Provided that human attentional resources are limited, Skehan believes that attention can only be allocated to certain aspects of performance to the detriment of others. Therefore, when task demands increase, learners first allocate attentional resources to the content of the task (i.e., fluency), and what remains is assigned to linguistic form (i.e., complexity and accuracy). If the content demands are extremely high, complexity and accuracy may compete for attention, and one may cause a negative impact on the other (e.g., Michel et al., 2019; Sample & Michel, 2015). Skehan's (2009, 2015) model is in accordance with Kormos' (1999, 2000) conceptualization of the role of attention in self-monitoring, in that both suggest that L2 production stages (i.e., conceptualization, formulation) may face a competition for cognitive resources, generating a potential trade-off between complexity and accuracy measures of L2 oral performance. Additionally, Kormos postulated that attentional limitations could limit the number and type of errors (e.g., lexis, grammatical, phonetic) noticed by the speaker and available for self-monitoring. Skehan suggested three factors contributing to the difficulty of the task, namely, code complexity, cognitive complexity and communicative stress, and other learner factors. Nevertheless, his model was unable to explain the phenomenon of dual-task performance and divided attention

nor was concerned with how tasks should be sequenced to promote L2 learning outside the foreign language classroom (Robinson, 2011).

An alternative strand of TBLT research attempting to manipulate learners' attention is the work within the cognition hypothesis (Robinson, 2011), grounded on information-processing theories, interactionist research, and psychological models such as Wickens' (1989) model of dual-task performance. Robinson's (2011) cognition hypothesis claims that learners can simultaneously access multiple and noncompetitive resource pools, and predicts that the increase of cognitive demands of a task is likely to direct attentional and memory resources to linguistic features and therefore lead to greater L2 grammatical and lexical accuracy and complexity, as long as learners draw from different pools of attentional resources. In order to identify specific task factors that should be manipulated to make tasks more or less cognitively demanding, the triadic componential framework (Robinson, 2001a; Robinson & Gilabert, 2007) distinguishes resource-directing from resource-dispersing dimensions. The former refers to those in which the demands on language use can be met by manipulating the manner in which the information is presented (e.g.,  $\pm$  few elements,  $\pm$  reasoning). In contrast, the latter refers to those that mirror the processing conditions under which real-time language is often used (e.g.,  $\pm$  planning time,  $\pm$  prior knowledge). Increasing task complexity along resource-directing dimensions may potentially direct cognitive resources to linguistic form, thus, leading to a greater accuracy and complexity in oral production, often at the expense of fluency (Gilabert, 2007; Ishikawa, 2008; Robinson, 2001b). In contrast, increasing task complexity along resource-dispersing dimensions could pose greater demands on attention and working memory, thus, depleting attention from the language code, which could be detrimental to L2 production.

Finally, increased task complexity often results in significantly higher ratings of task difficulty, mental effort, and anxiety while keeping task interest and motivation unaffected (Robinson, 2001b). Research has shown that task complexity manipulation may affect CAF measures differentially in speaking tasks. For example, Jackson and Suethanapornkul's (2013) systematic review found nonsignificant task complexity effects for *syntactic complexity* ( $d = -0.02$ ), small positive effects for *accuracy* ( $d = 0.28$ ); and a negligible but positive effect for *lexical complexity* ( $d = 0.03$ ), suggesting that increased task complexity led to larger lexical variety/diversity/density, at the expense of speaking *fluency* ( $d = -0.16$ ), consistent with the cognition hypothesis. However, the relation between task complexity and L2 pronunciation remains largely unexplored. TBLT research has previously assessed L2 pronunciation as part of speaking fluency or lexical accuracy (Kim & McDonough, 2008) or in terms of pronunciation errors (Kuiken & Vedder, 2011), but few studies have investigated it in relation to global dimensions of pronunciation (Gordon, 2021) or through acoustic analyses (Mora-Plaza, 2023; Solon *et al.*, 2017). However, to our current knowledge, no studies to date have investigated to what extent the predictions of the cognition hypothesis hold for L2 pronunciation in pronunciation-unfocused tasks.

### Task complexity and L2 pronunciation

One of the current discussions within the realm of L2 pronunciation instruction is whether task-based methodologies can promote attention to L1–L2 phonological differences and create opportunities for learners to acquire L2 sound contrasts and phonological features, and update the phonological form of their lexical representations. Reactive form-focused instructional techniques (e.g., negative feedback) and task

design and manipulation (e.g., modality, repetition, complexity) have been found to lead to more accurate L2 pronunciation during communicative task performance (Gurzynski-Weiss et al., 2017). For example, Solon et al.'s (2017) study revealed that L2 Spanish learners produced one out of five Spanish vowel monophthongs (/e/) with a more target-like quality (as assessed through acoustic analyses of formant frequencies) in the complex than the simple version of the task.

On the one hand, recent evidence suggests that, when tasks are designed to promote a focus-on-phonetic form (i.e., pronunciation-focused), task complexity positively impacts L2 pronunciation accuracy and subsequently leads to gains in L2 phonological development (e.g., Mora-Plaza, 2023). In the same vein, Mora-Plaza et al. (2018) and Mora-Plaza (2023) reported gains in the production of difficult English vowel contrasts for L1 Catalan learners, as measured through Euclidean and Mahalanobis distances, respectively, between L2 confusable vowels. Lastly, Gordon (2021) found beginner-level English as a foreign language (EFL) learners assigned to a complex-decision-making task condition intervention to outperform those assigned to a simple-decision-making task condition in comprehensibility (but not in accentedness) after treatment. Together, these studies provide evidence of the potential of task complexity to draw learners' attention to phonological form and improve pronunciation through a communicative form-focused intervention.

On the other hand, in pronunciation-unfocused tasks, it remains uncertain whether increasing task demands might have detrimental effects on L2 pronunciation. For example, Kuiken and Vedder (2011) investigated the influence of task complexity along  $\pm$  reasoning demands on L2 performance as a function of mode (i.e., written/oral) and L2 proficiency (i.e., high/low). Although this study was not intentionally designed to assess the impact of task complexity on L2 segmental and suprasegmental speech features in the oral version of the task, descriptively, increased task complexity was found to increase lexical and grammatical accuracy but to decrease pronunciation accuracy, especially in the case of low proficiency L2 learners. These findings point to the possibility of a potential trade-off between lexico-grammatical and pronunciation, where more attentional resources might be allocated to the control of lexical and grammatical form than phonetic form (Derwing, Munro & Wiebe, 1998), in line with Skehan's (2009, 2015) hypothesis. With a comparable cohort to the present study, Mora, Mora-Plaza & Bermejo Miranda's (2024) study revealed that learners produced significantly fewer lexico-grammatical errors in the complex version of a monologic oral task than the simple version (Gilabert, 2007; Robinson, 2001b), but the opposite pattern was found for pronunciation, with pronunciation errors being more frequent (though not to a significant extent) in the complex than the simple task. Increased attention to lexical and grammatical aspects during task performance may thus make it difficult for learners to allocate attentional resources to pronunciation. For instance, Crowther et al. (2018) assessed the extent to which comprehensibility and accentedness ratings were related to segmental and suprasegmental aspects of L2 speech in three tasks differing in cognitive complexity. One of the findings was that learners' speech was rated as significantly more strongly accented (but not less comprehensible) in the complex than the simple task, although the effect sizes were relatively small. These findings would lend support to Kormos' (1999, 2000) attention and monitoring model of speech processing postulating that the demands of the task might determine the number of attentional resources available during self-monitoring. Consequently, learners might need to pay more attention to lexical and grammatical aspects of speech to successfully convey the message than to segmental or suprasegmental

aspects. Kormos (1999, 2000) claimed that grammatical and lexical slips of the tongue (i.e., a measure of lexico-grammatical accuracy) are more likely to be detected at a different time than phonological errors while accounting for interindividual variation in L2 proficiency.

### Current study

To date, L2 acquisition and TBLT research has emphasized the well-established relation between task complexity and speech production and development. While most research studying the effects of increasing task demands has focused on learners' grammatical, lexical, and pragmatic performance and development, much less attention has been given to L2 pronunciation and prosody. Therefore, it is necessary to understand whether L2 pronunciation can be attended to when the demands of a task place great strain on learners' production processes, and learners need to invoke all linguistic resources available.

The primary aim of the current study was to test the predictions of the cognition hypothesis (Robinson, 2001a, 2011) on L2 pronunciation by manipulating task complexity (simple vs. complex) in a pronunciation-unfocused task. The choice of the cognition hypothesis and triadic componential framework (Robinson, 2001a, 2011) as the theoretical model for the conceptualization of this study was motivated by (1) the model's comprehensive account of the dimensions of task complexity that may influence L2 performance (i.e., the role of reasoning demands as a resource-directing variable, and planning time and prior knowledge as a resource-dispersing variable); (2) the model's predictions of task complexity regarding L2 oral development (as in Révész, 2009); and (3) the comparability with previous pronunciation-focused (e.g., Gordon, 2021; Mora-Plaza, 2023) and pronunciation-unfocused (e.g., Kuiken & Vedder, 2011) studies which methodologically manipulated task complexity along Robinson's triadic componential framework and theoretically explained their findings in light of the cognition hypothesis.

First, we assessed learners' L2 segmental accuracy (VOT in oral stops and degree of contrastiveness and nativelikeness in vowels) in a simple and complex version of a problem-solving task. Then we obtained NL' ratings of comprehensibility and accentedness on learners' speech sample excerpts from both task versions. Finally, in order to establish whether task effects were consistent across acoustic and global measures of pronunciation accuracy, and whether acoustically more accurate productions of oral stops (VOT) and vowels (Mahalanobis distances) predicted NL' ratings of comprehensibility and accentedness, we assessed the relationship between acoustic and global measures in the simple and complex tasks. Accordingly, we formulated the following research questions and hypotheses:

**RQ1:** How does task complexity affect learners' production of voice onset time (VOT) in word-initial stressed voiceless plosives (/p, t, k/)?

**RQ2:** Does task complexity have an effect on learners' vowel production (/i:/-/ɪ/, /æ:/-/ʌ/)?

**RQ3:** How does task complexity affect learners' ratings of comprehensibility and accentedness?

**RQ4:** To what extent are acoustic and global measures of learners' pronunciation related in the simple and the complex task?



In line with Kuiken and Vedder (2011) and Mora et al.'s (2024) studies, complexifying the task along resource-directing dimensions ( $\pm$  reasoning) in pronunciation-unfocused tasks is predicted to draw learners' attention away from phonological form, negatively affecting pronunciation accuracy (RQ1 and RQ2). This would lend support to Kormos (2000) and Skehan's (2009, 2015) predictions on potential trade-offs between areas of L2 oral performance, and would contradict Robinson's (2011) cognition hypothesis. In line with the findings by Crowther et al. (2018), accentedness but not comprehensibility ratings (RQ3) may be affected by task complexity. Acoustic and global measures are hypothesized to be associated (RQ4) in both simple and complex tasks (Crowther et al., 2018); especially foreign accent scores are expected to be moderately related to VOT productions (Riney & Takagi, 1999) and vowel accuracy (Munro, 1993).

## Method

### Participants

Eighty-two undergraduate advanced EFL learners (see Table 1 for demographics) participated in the study for course credit (*female* = 70, *male* = 12). They were Catalan-Spanish bilinguals who had learned English as an L2 mainly through formal instruction at school since the age of five. In this bilingual context, learners varied in Catalan-Spanish dominance (12 Catalan-dominant, 23 Spanish-dominant, 42 balanced bilinguals), but this was not expected to affect their L2 pronunciation performance significantly because Spanish and Catalan speakers do not differ in their use of short-lag stops and share the vowel categories /i/ and /a/, the only high-front and low-central vowels in their vocalic system. Participants were randomly assigned to two groups differing in task order: simple (S)>complex (C) ( $N = 45$ ) or C>S ( $N = 37$ ).

Thirteen NL (six males, seven females, *mean age* = 32.9, *SD* = 7.7) were recruited to evaluate the L2 learners' speech samples for comprehensibility and accentedness. They were experienced EFL teachers speaking either British (46%) or American (53%) English varieties. They reported being very familiar with Spanish/Catalan-accented

**Table 1.** Participants' demographics.

Measure	<i>M</i>	<i>SD</i>	95% CI	
			Lower	Upper
Age at testing (years)	20.23	2.94	19.59	20.88
Age of onset of L2 learning (years)	5.13	1.78	5.74	5.53
Spoken L2 input <sup>a</sup>	4.42	0.59	4.29	4.55
Spoken L2 output <sup>a</sup>	3.46	0.96	3.25	3.68
Written L2 input <sup>a</sup>	3.38	1.05	3.15	3.62
Written L2 output <sup>a</sup>	3.43	1.02	3.20	3.65
Self-estimated overall proficiency <sup>b</sup>	7.12	0.97	6.90	7.33
Self-estimated pronunciation proficiency <sup>c</sup>	6.43	1.46	6.11	6.75
Vocabulary size (0–10,000 words) <sup>d</sup>	6741	1232	6468	7013
L2 proficiency (score 0–120) <sup>e</sup>	96.60	13.40	93.68	99.53

<sup>a</sup>1 = never, 2 = yearly, 3 = monthly, 4 = weekly, 5 = daily.

<sup>b</sup>Reading, listening, speaking, writing: 9-point Likert scale from 1 = very poor to 9 = native-like.

<sup>c</sup>Pronunciation only: 9-point Likert scale from 1 = very poor to 9 = native-like.

<sup>d</sup>Measured by a Yes/No vocabulary size test (Meara & Miralpeix, 2015).

<sup>e</sup>Measured by an elicited imitation task (Wu, Tio & Ortega, 2021).

English on a 9-point Likert scale (1 = “not familiar at all”; 9 = “very familiar”;  $M = 8.5$ ,  $SD = 0.9$ ).

An additional group of eight native speakers (NS) of Southern British English (three males, five females) were recruited to perform the same speaking tasks as learners to obtain baseline speech data. They were all EFL teachers (*mean age* = 41.1,  $SD = 12.0$ ), who had had a predominantly monolingual upbringing and had lived in Spain for 16.3 years ( $SD = 12.6$ ).

### *L2 oral narrative: the dinner table task*

Learners performed a simple and a complex monologic version of the dinner table task (Ur, 1981) in a recording booth on different days. In this task, learners had to decide on and justify the seating arrangement of several characters at different tables. To eliminate the potential confound of task sequence and complexity, half of the participants performed the tasks in a S>C order and half in the C>S order. Both the simple and the complex versions of the task involved four stages: 1) a listening pretask, 2) a speaking task part 1, 3) a speaking task part 2, and 4) a posttask questionnaire.

The pretask consisted of a listening activity, in which participants were given an answer sheet containing a list of target words that defined each one of the characters (Appendix A). They were asked to read the words out loud and ask for the meaning of any they did not understand. Then, they heard a recorded description of the dinner party and were asked to match each character with the words that related to them. Based on Willis' (1996) task-based learning framework, the purpose of this priming stage was to familiarize learners with the task procedure, to introduce the characters' personality traits, professions, and hobbies, and to provide learners with the necessary linguistic resources (i.e., semantic and phonetic form of words) to be able to successfully complete the main communicative task. In this way, we attempted to reduce cognitive demands at the level of resource-dispersing variables (Robinson, 2011).

In the first part of the speaking task, participants were given a picture of tables with six characters sitting at them (two characters at three tables in the simple version, three characters at two tables in the complex version), and they were asked to carefully consider the seating arrangement and justify why it would not work based on the attendees' personality traits, professions, and hobbies (Appendix B). They were given 1.5 minutes of planning time and were encouraged to provide as many reasons as they could think of by exploiting all personality features while considering an appropriate seating arrangement to ensure a comprehensible and thoughtful response. The same procedures were applied to the second part of the task, except that the tables were empty and participants were given six cards (one for each character) and were asked to decide on a new seating arrangement that would lead to a pleasant party. The task was closed and guided as it was key for assessment purposes that participants produced specific items containing the target vowels and consonants. However, participants were completely unaware that the task was designed to analyze their pronunciation and their attention was not drawn to the target phonetic forms at any time, hence, it was a purely pronunciation-unfocused task. The dinner table task, including its instructions and printable materials, is deposited in the open science repository, SLA Speech Tools (Mora-Plaza, Saito, Suzukida, Dewaele & Tierney, 2022: <http://sla-speech-tools.com/>).

Finally, a task-performance questionnaire was administered immediately after learners had completed the task (Appendix C). They were asked to rate how well they



had performed in the task, how difficult they had perceived the task to be, how much mental effort they had put into it, and how anxious they had felt during task performance on a 9-point scale (1 = *very poorly, not difficult, no mental effort, not anxious*; 9 = *very well, extremely difficult, extreme mental effort, very anxious*).

### Target Words

The target words comprised names (e.g., *Tilly Killey, John Butler*) and adjectives (e.g., *impulsive, deceitful*) describing the characters' personality, beliefs, occupation, and interests, and were identical in the simple and complex versions of the task, except for some characters' names and surnames, which contained the same target vowels, but had to be different to identify different characters and their pictures. Both versions contained 27 words with the target consonants (10 /p/, 9 /t/, 8 /k/) and 36 words with the target vowels (/i:/, /ɪ/, /æ/, and /ʌ/; 9 each) (Appendix D). Obtaining a minimal number of productions of the target sounds from the L2 learners was a prerequisite for obtaining reliable acoustic measures.

VOT was chosen as an index of segmental pronunciation accuracy for consonants because it has been found to be sensitive to experience-related factors in L2 speech production for Catalan/Spanish learners of English (Gorba & Cebrian, 2021), and to linguistic environment (Olson, 2020) and local contextual effects in code-switching tasks (Olson, 2013). While English has long-lag voiceless stops (40–80 ms; e.g., Docherty, 1990), they are short lag in Spanish (7–20 milliseconds; e.g., Castañeda, 1986). Due to a lack of awareness of cross-language differences between L1 and L2 voiceless plosives (Flege, 1995) and reliance on different phonetic cues (e.g., presence or absence of closure voicing; Mora, Rochdi & Kivistö-de Souza, 2014), Spanish learners of English tend to produce English stops with “intermediate” VOT values falling short of English VOT.

The target vowels (/i:/, /ɪ/, /æ/, /ʌ/) form high and low phonologically contrastive vowel pairs in English (/i:/-/ɪ/ and /æ/-/ʌ/, respectively) that are difficult to distinguish qualitatively in perception and production for L1–Catalan/Spanish learners of English. This is because the vowels in each pair are perceptually mapped onto a single native vowel category that is acoustically located between the English vowels (Spanish /i/ for the high vowel contrast and Spanish /a/ for the low vowel contrast). Cross-language perceptual assimilation tasks (Cebrian, 2019) show that English /i:/ and /ɪ/ are identified as Spanish /i/, whereas English /æ/ and /ʌ/ are identified as Spanish /a/ (Rallo Fabra & Romero, 2012). In addition, Spanish learners of English produce very small spectral distances in the production of these L2 vowel contrasts (Darcy, Mora, & Daidone, 2016), failing to distinguish contrastive vowels effectively in production. Although the high vowels are less consistently identified as Spanish /i/ than the low vowels are as Spanish /a/, both are confusable vowel contrasts posing great difficulty in perception and production at the phonetic (prelexical) and lexical levels for advanced Spanish learners of English (Mora & Mora-Plaza, 2019). Although Spanish learners of English have been shown to rely on temporal cues (i.e., duration) in the perception and production of /æ/-/ʌ/ and /i:/-/ɪ/ (Cebrian, 2006; Mora & Fullana, 2007; Rallo-Fabra & Romero, 2012), we opted for focusing on spectral rather than temporal cues, as recent research has shown that duration ratios for /i:/-/ɪ/ in production are not larger than those of NS (Cebrian, Gorba & Gavalda, 2021), and because temporal cues are likely to be more readily affected by speakers' individual differences in speaking style (e.g., speech rate) and durational variability associated with the position of the target words

in the utterance and their prosodic prominence. Therefore, potential task complexity effects on vowel production accuracy were examined only with respect to the spectral aspects of the target vowel contrasts.

### **Manipulation of task complexity**

Task complexity was manipulated along  $\pm$  reasoning demands (Robinson, 2011) by varying the number of characters seated at each table (two vs. three), and the combination of personality traits (coherent vs. incoherent). The complex task was therefore more demanding than its simple counterpart because sitting three people at the same table with incoherent traits requires more cognitively demanding decisions. Therefore, the two versions of the task were identical in terms of several elements and target lexical items but differed in the characters' names and the distribution of table and personality characteristics. Manipulation of task complexity in the present study differs from previous studies in that participants were given the target words they had to use both in the simple and complex versions of the task.

The outcome of the task-performance questionnaires revealed that overall learners perceived to have performed worse, feeling more anxious and employing more effort in the complex than the simple task. Learners perceived the complex task to be significantly more difficult ( $M = 5.11$ ,  $SD = 1.82$ ) than the simple task ( $M = 4.63$ ,  $SD = 1.88$ ;  $F[81] = 4.19$ ,  $p = .044$ ) and to require significantly more mental effort ( $M = 5.93$ ,  $SD = 1.76$ ) than the simple task ( $M = 5.41$ ,  $SD = 1.73$ ;  $F[81] = 9.40$ ,  $p = .003$ ), which is in line with Robinson's hypothesis and empirical findings (Robinson, 2001b).

### **Procedures**

Participants first filled in an online background questionnaire. In the lab, they performed an elicited imitation task (Wu *et al.*, 2021), a Yes/No vocabulary size test (Meara & Miralpeix, 2015), and the L2 oral production task, immediately followed by the posttask questionnaire. The simple and complex versions of the speaking task were performed in two different days and counterbalanced in order to mitigate potential task interference and carryover effects, and minimize fatigue effects on participants, ensuring each version received undivided attention and engagement. Learners' oral productions were recorded in a soundproof booth on Marantz PMD661 solid-state digital recorders with an external Shure SM58 voice microphone at a sampling frequency of 44.1 KHz.

### **Data analyses and speech production measures**

The speaking task generated approximately 5 minutes of speech per task and participant. The average length of oral narratives was similar in both versions of the task for learners (Simple:  $M [sec] = 309$ ,  $SD = 104$ ,  $M [min] = 5.16$ ; Complex:  $M [sec] = 308$ ,  $SD = 108$ ,  $M [min] = 5.13$ ) and NS (Simple:  $M [sec] = 320$ ,  $SD = 102$ ,  $M [min] = 5.33$ ; Complex:  $M [sec] = 334$ ,  $SD = 132$ ,  $M [min] = 5.56$ ).

Recorded amplitude-by-time waveforms were automatically segmented into speech or pause ( $> 250$  milliseconds) intervals using the *Annotate to TextGrid* (silences) command in Praat (Boersma & Weenik, 2015), manually adjusted for segmentation inaccuracies, and orthographically transcribed. Filled and silent pauses, analysis of speech (AS) units, speech dysfluencies (repetitions), and lexical, grammatical, and

pronunciation errors were manually annotated (Appendix E). The transcribed text and the corresponding audio files were submitted to the WebMAUS Basic automatic segmentation system (Schiel, 1999) to obtain labeled word and sound intervals.<sup>1</sup> Words containing the target oral stops and vowels were identified for acoustic measurement.

The VOT in the prevocalic voiceless oral stops (/p/, /t/, /k/) of word-initial stressed syllables in the target words was annotated manually in Praat and measured in milliseconds from the onset of the release burst of the stop consonant to the first positive peak of periodic energy of the following vowel. To exclude potential measurement errors, VOT durations outside 2.5 SD from each subject's mean were screened (1.4%).

Vowel quality was measured by extracting frequency measurements (F1, F2) from a 10ms window by manually placing the cursor at the midpoint of the steady-state portion of the target stressed vowels. Frequency values were then converted to Bark (B), a psychoacoustic scale measure that changes frequency differences in vowel quality in terms of their impact on human perception and helps minimize interspeaker variability in vocal tract size. Bark-converted frequencies were then used to estimate the degree of vowel height (B1) and frontness (B2) and to determine distributions of vowel tokens for the English vowel categories /i:/, /ɪ/, /æ/ and /ʌ/ on a two-dimensional B1–B2 space. Changes in vowel production accuracy resulting from task complexity were estimated through Mahalanobis distances between learners' vowel productions and the corresponding vowel spaces of the control NS. Mahalanobis distances compute the distance in standard deviations between a point and the centroid of the distribution and take into consideration not only the centroid location but also the spread and orientation of the reference distribution, thus reflecting token variability (Kartushina, Hervais-Adelman, Frauenfelder & Golestani, 2015; Melnik-Leroy, Turnbull & Peperkamp, 2022). We computed Mahalanobis distance scores (DS) between vowels for the high (/i:/-/ɪ/) and the low (/æ/-/ʌ/) vowel contrasts in the simple and the complex task as a measure of *contrastiveness* (i.e., how distinct vowel quality was within the contrast), hence, a larger distance meant less of an overlap between the two vowels. We also computed, for both tasks, Mahalanobis distances between learners' and NS' productions of these vowels (/i:/, /ɪ/, /æ/, /ʌ/) as a measure of *nativelikeness* (i.e., how much learners' vowel qualities approximate those of NS), hence, a smaller distance meant a more target-like production. Although increased contrastiveness between vowels is, by hypothesis, assumed to index improved accuracy in production in phonetic training studies using minimal-pair testing stimuli (e.g., Melnik-Leroy et al., 2022), the spontaneous nature of the oral task we used to elicit L2 speech did not allow us to have control over the stimuli the learners produced. Thus, a measure of nativelikeness separately computed for each target vowel was deemed more appropriate than a measure of contrastiveness to gauge task complexity effects on vowel production accuracy within subjects.

Measures of comprehensibility and accentedness were obtained from 13 NL who rated 164 speech excerpts approximately 45-sec long ( $M = 46.2$ ,  $SD = 2.6$ ) from the second part of the learners' simple and complex versions of the speaking task on 9-point scales ( $1 = \textit{very difficult to understand / not accented at all}$ ,  $9 = \textit{very easy to understand / very strongly accented}$ ). We extracted the excerpts from the second part of the participants' performance owing to heightened cognitive demands. In this part, they had to propose a new seating arrangement that would guarantee a pleasant party after having

<sup>1</sup>The link to WebMAUS Basic website is <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>

provided reasons why the given arrangement would not work. The speech samples had been previously normalized for peak and mean amplitude and bandstop-filtered at 50Hz. Each rater judged each speech sample twice, first for comprehensibility and then for accentedness. These dimensions and the rating procedure were explained to raters in a training session. The rating task contained three practice trials from participants not included in the study. The 164 speech samples from the 82 learners were distributed randomly in four rating sessions taking place on different days. Within every session, lasting approximately 1 hour, all the speech samples were fully randomized and presented in blocks of 15 separated by short breaks in order to minimize the influence of familiarity. Interrater reliability (Cronbach's alpha intraclass correlation coefficients) of the NL' ratings was high for comprehensibility ( $\alpha = .90$ ) and accentedness ( $\alpha = .94$ ), so single mean scores per speech sample were computed by averaging across all ratings for each rated measure. No notable differences were detected between British and American raters in their evaluations of the speech samples.

### Statistical Analyses

Fixed-effects structures were defined for each one of the models. Fixed and random-effects structures for all analyses in this study were selected based on the best fitting model (i.e., comparing Akaike information criterion [AIC] estimators across models), and random slopes were only included if they improved the model's fit (i.e., AIC decreased), provided that the model could converge. Finally, Bonferroni adjustments were used for pairwise contrasts, and parameter estimates are reported in [Appendix F](#). The assumptions of collinearity, normal distribution of residuals, and homoscedasticity were all met.

## Results

Although we were expecting, in light of previous research (Jackson & Suethanapornkul, 2013), the simple and the complex versions of the dinner table task to affect speech production linguistically in terms of lexical and grammatical accuracy and complexity, as well as in speaking fluency, we found the task manipulation effects to be very small in magnitude (see [Table 2](#) below). The results indicated large interlearner variability in all the measures and very small effects (if at all observable) for lexical and grammatical

**Table 2.** Mean complexity (C), accuracy (A), and fluency (F) scores in the simple and complex tasks.

Measures	Simple		Complex	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
C AS Unit length (words)	13.2	3.5	13.5	2.6
C Guiraud's index	7.0	0.8	6.9	0.8
A Lexico-grammatical errors ( $\times 100$ words)	2.0	1.3	2.1	1.3
F Speech rate (syllables $\times$ min)	136	23	132	24
F Mean syllable length (milliseconds)	311	35	319	64
F Internal (AS unit) pauses $\times$ min	26.7	13.2	27.2	8.9
F External (AS unit) pauses $\times$ min	13.3	8.4	12.3	3.6
F Internal mean pause duration (milliseconds)	597	90	601	114
F External mean pause duration (milliseconds)	691	207	764	271
F Repetitions and self-corrections ( $\times 100$ words)	1.6	1.02	1.6	0.87
F Mean length of run (syllables between pauses $> 250$ milliseconds)	5.9	1.7	5.9	1.6

complexity and accuracy, as well as for measures of speed and repair fluency. For breakdown fluency, the duration of pauses between AS units was substantially longer in the complex than in the simple version of the task.

A series of mixed-effects models were run in SPSS 27 on all the complexity, accuracy, and fluency measures, which included a random intercept for *subject* and *task sequence* (S>C, C>S) as a fixed factor in addition to *task* (simple, complex) to control for potential effects of learners performing the simple or the complex task first. The analysis did not reveal any significant effects of *task*, except for the duration of pauses at AS unit boundaries, which turned out to be significantly longer in the complex than the simple task ( $F[1, 161] = 12.98, p < .001$ ), probably indicating greater conceptualization and formulation difficulties in the complex than the simple task. The effect of *task sequence* did not reach significance in any of these analyses. Thus, our task complexity manipulation affected breakdown fluency significantly, but not other dimensions of oral production (i.e., complexity, accuracy). The need to elicit specific forms for acoustic measurement, which forced us to constrain the task in terms of lexical choice, probably limited the variability in learners' use of lexical and grammatical resources. However, as reported above, learners perceived the complex task as significantly more demanding, requiring greater effort, and posing higher levels of difficulty and anxiety compared to the simple task. Thus, we interpreted the significant effects of task complexity in breakdown fluency and learners' post-performance self-reports to suggest that the two versions of the task could vary in how much attention to phonetic form they allowed, which could potentially lead to differences in pronunciation accuracy.

### **Task complexity and consonant production**

Prior to assessing the effects of task complexity on learners' VOT productions (RQ1), we looked into the VOT differences between learners' and English NS' productions of /p/, /t/, and /k/. Applying a square root transformation to VOT values, linear mixed-effects models with *speaker group* (NS, learner), *consonant* (/p/,/t/,/k/) and their interactions as predictors, and by-*subject* random intercepts, revealed significant main effects of *speaker group* ( $F[10985] = 33.11, p < .001$ ), *consonant* ( $F[10985] = 726.27, p < .001$ ) and a significant *speaker group*  $\times$  *consonant* interaction ( $F[10985] = 55.52, p < .001$ ). As expected, these results showed that NS produced significantly more aspirated consonants ( $M = 63.25$  ms,  $SD = 21.61$ ) than L2 learners ( $M = 42.21$  ms,  $SD = 19.46$ ) did (i.e., 49.85% longer VOT) and that /p/ was the least aspirated consonant, followed by /k/ and /t/. The VOT in the production of English /t/ tends to be more target-like in Spanish learners of English than /p/ and /k/ because learning to produce a different L2-specific place of articulation (alveolar in English vs. dental in Spanish) enhances overall articulatory accuracy in the degree of stricture and laryngeal timing (e.g., Mora, 2008). The interaction arose because, although VOT differences were present in all target consonants, NS produced greater VOT values for /k/ > /t/ > /p/ and learners for /t/ > /k/ > /p/ (Table 3).

The effects of task complexity on learners' VOT productions were assessed by fitting the learners' VOT to a linear mixed-effects model with a gamma regression function with *task* (simple, complex), *consonant* (/p/, /t/, /k/) and their interactions as fixed effects. *Task sequence* (S>C, C>S) was also included as a fixed effect to control for potential task order. The random-effects structure included a random intercept for *subject* (see Appendix F for parameter estimates). The model yielded main effects of

**Table 3.** Voice onset time (in milliseconds) for NS' and learners' productions of oral stops in initial stressed position.

	NS				Learners			
	M	SD	95% CI		M	SD	95% CI	
			Lower	Upper			Lower	Upper
/p/	49.77	20.56	47.42	52.13	25.39	16.13	24.81	25.97
/t/	68.67	20.38	66.66	70.67	54.52	20.97	53.84	55.20
/k/	71.32	23.89	68.75	73.90	46.73	21.28	46.00	47.46

Note: *M*: mean, *SD*: standard deviation, *CI*: confidence interval.

task ( $F[9956] = 26.08, p < .001$ ) because learners' VOT productions were significantly more aspirated (i.e., 3.46% more accurate) in the simple ( $M = 43.99\text{ms}, SD = 23.34, 95\% \text{ CI} = 43.33\text{--}44.65$ ) than the complex ( $M = 42.52\text{ms}, SD = 22.97, 95\% \text{ CI} = 41.89\text{--}43.15$ ) task, and a significant main effect of *consonant* ( $F[9956] = 2907.01, p < .001$ ) and *task sequence* ( $F[9956] = 7.02, p = .008$ ). The *task*  $\times$  *consonant* interaction did not reach significance ( $F[9956] = .86, p = .423$ ). Bonferroni-adjusted pairwise contrasts indicated that the main effect of *task* was driven by the three target consonants: /k/ ( $t[9956] = 4.02, p < .001$ ), /t/ ( $t[9956] = 2.57, p = .010$ ) and /p/ ( $t[9956] = 2.29, p = .022$ ) (Figure 1). The main effects of *task* were not significant ( $F[1021] = 2.30, p = .129$ ) when the same model structure was applied to NS' VOT productions.

### Task complexity and vowel production

#### Task complexity effects on vowel distribution

Preliminary analyses of vowel quality unexpectedly revealed changes in NS' vowel quality in the same direction as those observed in learners, which we attributed to the unbalanced distribution of the /æ/ and /ʌ/ vowel tokens across words and the use of

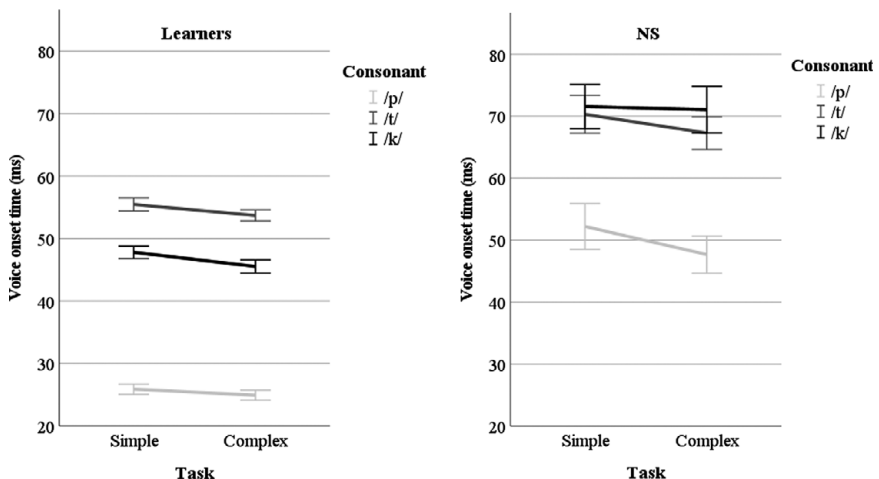


Figure 1. VOT (in milliseconds) as a function of speaker group, task, and consonant (error bars = 95% CI).



different proper names in the simple and the complex task<sup>2</sup>. Therefore, we decided to exclude all words corresponding to character names that were different in the simple and the complex version of the task (/i:/: *Keane, Keith*; /ɪ/: *Killey, Pickett*; /æ/: *Ann, Kang, Sam, Tang*; /ʌ/: *Butler, Cutler*). Given the high frequency of these words, this screening procedure resulted in considerable data loss, as a further group of 17 learners had to be excluded for not meeting the criterion of having at least three tokens of each one of the four target vowels. Consequently, the final vowel data set consisted of a total of 5,426 vowel tokens for learners ( $N = 60$ ) distributed relatively evenly by vowel contrast (/i:/: 1494; /ɪ/: 1668; /æ/: 1160; /ʌ/: 1134) and 814 vowel tokens for NS ( $N = 8$ ) (/i:/: 225; /ɪ/: 274; /æ/: 171; /ʌ/: 144).

These vowel data (see Table 4) showed that learners produced a lower /i:/ (higher B1) and a higher /ɪ/ (lower B1) than NS did both in the simple and complex task, whereas the learners' /æ/ and /ʌ/ differed only minimally from NS' /æ/ and /ʌ/ in the simple task, but in the complex task /ʌ/ was less target-like (higher B1, i.e., lower more /a/-like articulation). As regards fronting (B2), learners realized the lax vowels /ɪ/ and /ʌ/ with higher B2 both in the simple and the complex task than NS did, indicating a more /i:/-like production of /ɪ/, and a more /æ/-like production of /ʌ/. Such learner-NS differences indicate less target-like vowel quality in the production of the English lax vowels /ɪ/ and /ʌ/ than in the production of the English tense vowels /i:/ and /æ/, which is consistent with English /i:/ and /æ/ being a closer match to the corresponding high /i/ and low /a/ Spanish vowel categories in perception and production. This is reflected in Figure 2, where the red ovals representing the distribution of learners' /i:/ in the simple (S) and complex (C) tasks overlap more largely with NS' distribution of /i:/ tokens (black dotted line) than the green ovals representing learners' /ɪ/ tokens overlap with the distribution of NS' /ɪ/ tokens. A similar picture can be observed for /æ/ and /ʌ/. It is also clear from Figure 2 that (a) NS' productions of /i:/ and /ɪ/ and of /æ/ and /ʌ/ result in truly contrastive nonoverlapping distributions (black dotted lines), whereas for learners the distributions of contrastive high and low vowels overlap considerably (colored ovals), and (b) there is very little difference between learners' vowel productions in the simple (solid-colored lines) and the complex task (dashed colored lines). We next present the analyses of Mahalanobis distances between contrastive vowels

<sup>2</sup>It was only for the low vowels /æ/ and /ʌ/ that substantial differences in learners' vowel quality could be found between tasks. The articulation of /æ/ was fronter (higher B2) in the simple than the complex task, whereas the articulation /ʌ/ was less front, suggesting a less-target like production of /æ/ and more target-like production of /ʌ/ in the complex than the simple task. However, the same task complexity effects were found for NS, whose vowel qualities were not expected to change as a function of task complexity. This suggested that the task complexity effects in both learners and NS affecting the low vowels /æ/ and /ʌ/ could be an artefact of the unbalanced distribution of the /æ/ and /ʌ/ vowel tokens across words. Learners happened to produce 10.2% more /æ/ tokens and 10.7% more /ʌ/ tokens in the complex than the simple task, but crucially, some of the most frequent target words corresponding to character names for /æ/ and /ʌ/ in the simple task (/æ/: *Sam, Tang*; /ʌ/: *Cutler*) were different from some of the target words in the complex task (/æ/: *Ann, Kang*; /ʌ/: *Butler*) and unmatched for consonantal place of articulation. Close inspection of the B2 differences between the words in the simple and the complex task suggested that the observed task complexity differences were mainly driven by the B2 differences between the different words in the simple (*Sam*: 11.38/11.50, *Tang*: 11.56/11.58, *Cutler*: 11.09/10.58) and the complex task (*Ann*: 11.67/11.71, *Kang*: 11.72/11.76, *Butler*: 10.65/10.16). As differences in articulation rate might have led to less peripheral vowel articulations (Deterding, 1997), we assessed potential task complexity effects on mean syllable duration, which did not reveal any significant differences for either learners ( $S = 311\text{ms}$ ,  $C = 317\text{ms}$ ;  $t(152) = -.771$ ,  $p = .442$ ) or NS ( $S = 280\text{ms}$ ,  $C = 288\text{ms}$ ;  $t(30) = -.521$ ,  $p = .610$ ).

**Table 4.** Descriptives of Bark-converted frequency measures by speaker group (learners, NS).

		Simple (B1)					Complex (B1)				
		N	M	SD	Min.	Max.	N	M	SD	Min.	Max.
Learners	/i:/	749	3.56	0.61	1.92	5.73	745	3.61	0.65	1.63	5.98
NS	/i:/	117	3.32	0.66	1.85	5.24	108	3.33	0.56	1.83	4.43
Learners	/ɪ/	720	3.82	0.59	2.20	7.29	948	3.85	0.56	2.37	5.61
NS	/ɪ/	124	4.09	0.57	2.54	5.55	150	3.95	0.52	2.68	5.62
Learners	/æ/	575	7.04	0.88	3.43	9.44	585	7.15	0.89	4.17	9.99
NS	/æ/	94	7.18	1.25	3.71	10.91	77	7.36	1.18	5.22	10.55
Learners	/ʌ/	565	6.15	1.37	1.89	10.12	569	6.06	1.35	1.81	9.29
NS	/ʌ/	72	6.09	1.08	3.54	8.16	72	5.97	1.07	3.52	8.64

		Simple (B2)					Complex (B2)				
		N	M	SD	Min.	Max.	N	M	SD	Min.	Max.
Learners	/i:/	749	14.25	0.97	7.53	16.33	745	14.24	0.92	9.48	16.18
NS	/i:/	117	14.38	0.93	12.69	15.82	108	14.41	0.93	12.87	15.90
Learners	/ɪ/	720	13.89	0.87	11.11	15.83	948	13.84	0.81	11.00	15.88
NS	/ɪ/	124	13.38	0.86	11.57	15.34	150	13.24	0.87	11.84	15.17
Learners	/æ/	575	11.47	0.83	8.57	13.85	585	11.57	0.90	7.90	14.40
NS	/æ/	94	11.50	0.61	9.95	12.70	77	11.66	0.62	10.14	12.86
Learners	/ʌ/	565	10.67	1.08	7.55	14.12	569	10.69	1.04	6.49	12.76
NS	/ʌ/	72	10.27	0.84	8.42	12.04	72	10.35	0.88	7.41	12.61

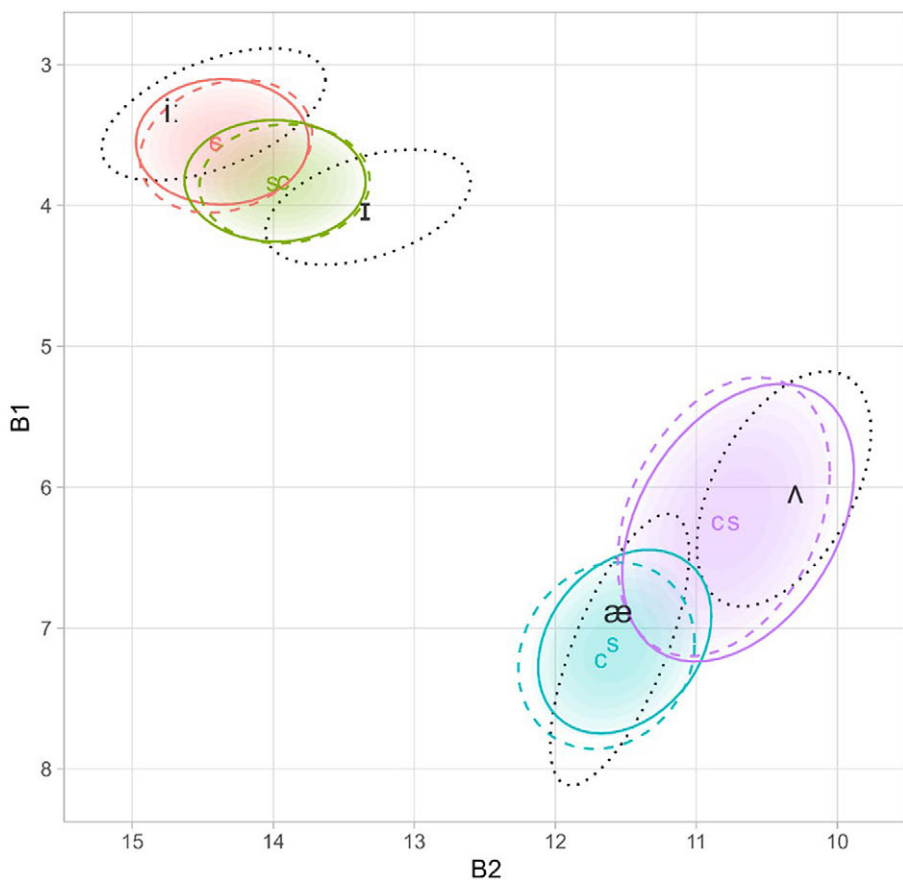
Note: Data from 60 learners and 8 NS (17 learners were excluded because they had less than 3 realizations of the target vowels) after screening words that were different in the simple and the complex task (*Ann, Butler, Cutler, Kang, Keane, Keith, Killely, Pickett, Sam, Tang*).

(contrastiveness) for learners and NS, and between learners and NS (nativeness) for each of the four target vowels.

The spontaneous nature of the speaking task generated large variability as regards the number of vowel tokens each participant produced and their quality (as assessed through first and second formant frequency measurements), which led to large dispersion clouds both in learners and NS. Still, the distribution of NS' vowel productions shows much less overlap between /i:/ and /ɪ/ and between /æ/ and /ʌ/, than the distribution of learners' vowel productions (Figure 2), indicating a larger degree of *contrastiveness* in NS than in learners, as expected. In addition, as shown in Figure 3, task complexity had very little effect overall on vowel quality and the distributions of learners' vowels do not seem to consistently present a larger overlap with NS' distributions for either the simple (S) or the complex (C) task, indicating little effect of task complexity on the degree of *nativeness* of the target vowels.

#### Task complexity effects on vowel contrastiveness

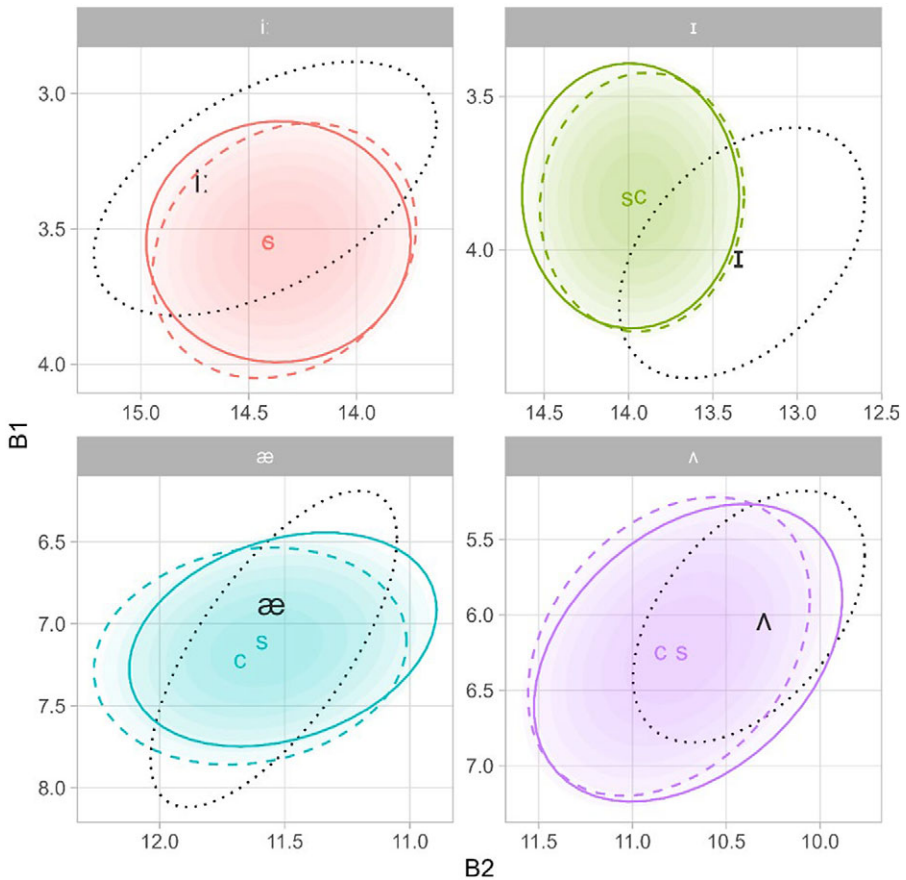
For contrastiveness, the log-transformed Mahalanobis DS between the vowel spaces of the vowel pairs /i:/-/ɪ/ and /æ/-/ʌ/ were submitted to fixed-effects models (separately for learners and NS) with *task*, *contrast*, and their interaction as fixed factors, and a random intercept for *subject*. We also included *sequence* (S>C, C>S) to control for potential task order effects (see parameter estimates in Appendix F). Overall, as expected (see Figure 4), the magnitude of the distinction learners made between contrastive vowels (between 4–10 SD) was much smaller than the distinction NS made (20–25 SD). For learners, tests of fixed effects revealed a main effect of *contrast* ( $F[1, 5430] = 380.7, p < .001$ ) and a significant *task* × *contrast* interaction ( $F[1, 5430] = 7.5, p = .006$ ), but the main effect of *task* did not reach significance ( $F[1, 5430] = 8.2, p = .121$ ). This



**Figure 2.** Distributions of Bark-converted first (B1) and second (B2) formant frequency values of the target vowels /i:/, /ɪ/, /æ/ and /ʌ/ as produced by learners (colored shaded ovals) and NS (dotted black lines) in the simple (S, solid lines) and complex (C, dashed lines) tasks. Ovals represent 32% confidence intervals.<sup>3</sup>

interaction arose because the log-transformed DS between the low vowels /æ/ and /ʌ/ were significantly larger in the complex (8.83) than the simple task (7.17;  $t[5430] = 2.82$ ,  $p = .005$ ), whereas the DS between the high vowels /i:/ and /ɪ/ did not differ significantly across tasks (3.75 vs. 3.76;  $t[5430] = 2.82$ ,  $p = .005$ ). The main effect of *contrast*, caused by log-transformed DS being much larger in low (7.87) than in high (3.76) vowels, was mainly driven by the DS between /æ/ and /ʌ/ in the complex task. It is uncertain why the more complex task led to larger contrastiveness for the low vowels but not for the high vowels as if the more complex task generated higher attention to phonetic form; we

<sup>3</sup>Given the huge variability in the spontaneous speaking data we collected where speakers contributed with different numbers of vowel tokens in different words and different sentence contexts and prosodic environments, we plotted the distribution of vowel clouds to a % confidence intervals (CI) that would show largely non-overlapping clouds for NS (32% CI). This allowed us to visualize the amount of overlap between the vowel categories for L2 learners (all of them showing large overlap) compared to the degree of overlap of the vowel clouds of non-native speakers (largely non-overlapping).



**Figure 3.** Distributions of Bark-converted first (B1) and second (B2) formant frequency values of the target vowels /i:/, /ɪ/, /æ/ and /ʌ/ as produced by learners (colored shaded ovals) and NS (dotted black lines) in the simple (S, solid lines) and complex (C, dashed lines) tasks. Ovals represent 32% confidence intervals.

would expect the effect to be observable also in the high vowels. Running the same mixed-effects model on the NS' data can help us elucidate this, as we would not expect the quality of their vowels and their degree of contrastiveness to change significantly as a function of task complexity. Unexpectedly, for NS main effects of *task* ( $F[1, 783] = 4.33, p = .038$ ) and *contrast* ( $F[1, 783] = 7.43, p = .007$ ) were found to be significant, whereas the *task*  $\times$  *contrast* interaction did not reach significance ( $F[1, 783] = .005, p = .994$ ). However, unlike learners, for NS, DS was larger overall for the high (26.05) than the low vowels (19.13). For NS, the main effect of *contrast* was driven by the joint contribution of DS being larger in high than low vowels both in the simple (25.21 vs 22.34;  $t[783] = 2.02, p = .044$ ) and the complex (27.16 vs. 18.08;  $t[783] = 1.84, p = .066$ ) tasks, whereas the main effect of *task* was driven by the joint contribution of DS being larger in the complex than the simple task (though nonsignificantly) in high vowels (27.15 vs 25.21;  $t[783] = 1.62, p = .106$ ) and DS being larger in the simple than the complex task (though nonsignificantly) in low vowels (22.34 vs. 18.08;  $t[783] = 1.38, p = .169$ ). The main effect of *sequence* did not reach significance for either learners ( $F$

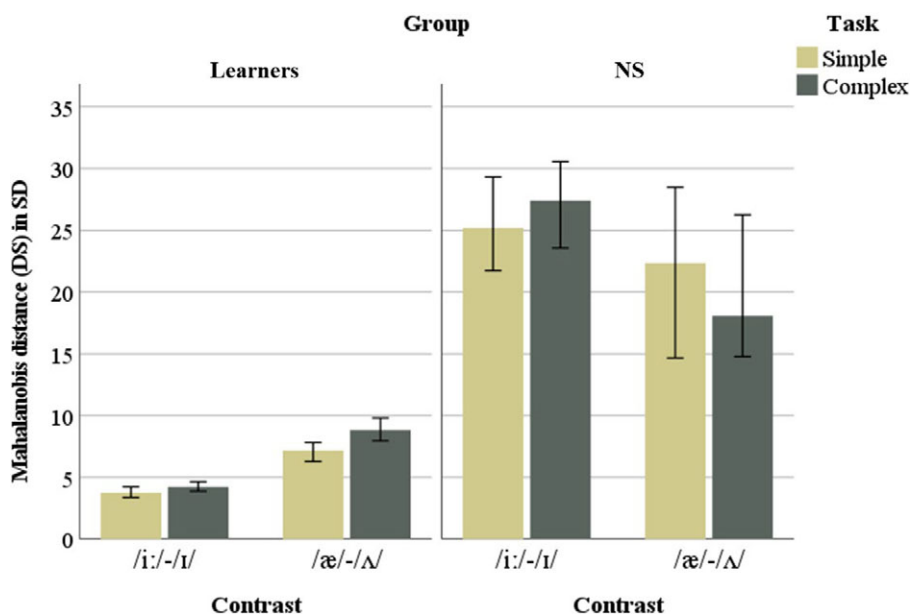
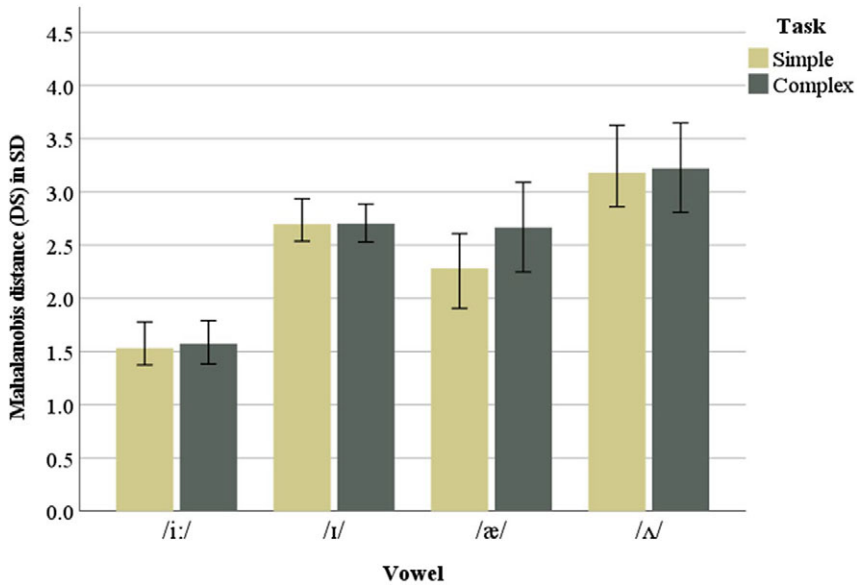


Figure 4. Median Mahalanobis distances between contrastive vowel pairs as a function of task, speaker group, and contrast (error bars = 95% CI).

[1, 5430] = .024,  $p = .876$ ) or NS ( $F[1, 783] = .005, p = .994$ ). As argued above, given the unbalanced contribution of vowel tokens across participants, we deemed Mahalanobis DS between contrastive vowel pairs to represent potential task complexity effects less reliably than Mahalanobis DS between learners' and NS' realizations of vowels embedded in the same set of target words even if, given the spontaneous nature of the speaking task, different speakers contributed a different proportion of target words to the final data set.

#### Task complexity effects on vowel nativelikeness

Mahalanobis DS between learners' productions and NS' vowel spaces (see Figure 5) do not show consistent task complexity effects on how native-like the vowel productions were. It was only for /æ/ that learners produced slightly larger DS with respect to NS, indicating a less target-like realization of /æ/, in the complex than the simple task. Mahalanobis DS (log-transformed) between learners' vowel productions and the vowel spaces of the native vowels /i:/, /ɪ/, /æ/, and /ʌ/ were submitted to a fixed-effects model with *task*, *vowel*, and their interaction, as well as *task sequence* as fixed factors and a random intercept for *subject*. Tests of fixed effects revealed a main effect of *vowel* ( $F[3, 5480] = 71.35, p < .001$ ) because in both tasks /i:/ (S: 1.53, C: 1.54) was realized with smaller median DS (more accurately) than /ɪ/ (S: 2.70, C: 2.70), and /æ/ (S: 2.28, C: 2.66) was realized in a more target-like manner than /ʌ/ (S: 3.18, C: 3.22). However, neither the effect of *task* ( $F[1, 5325] = .493, p = .483$ ) nor the *task* × *vowel* interaction ( $F[1, 5325] = 1.98, p = .102$ ) nor the effect of *sequence* ( $F[1, 5480] = 1.61, p = .205$ ) reached significance. According to Bonferroni-adjusted pairwise contrast tests, the less accurate realization of /æ/ we observed in the complex (DS: 2.66) than the simple task (DS: 2.28) did not reach significance either ( $t[5480] = -1.84, SE = .032, p = .066$ ) (see Appendix F).



**Figure 5.** Median Mahalanobis distances between learners' vowel productions and NS' vowel spaces as a function of task and vowel (error bars = 95% CI).

*Task complexity and global pronunciation ratings*

Task complexity effects on comprehensibility and accentedness were assessed through linear mixed-effects models with *task* and *task sequence* as predictors of NL' ratings, and by-*subject* and by-*rater* random intercepts. As shown in Table 5, learners' speech was rated as less comprehensible ( $F[2129] = 3.72, p = .054$ ), albeit nonsignificantly, and significantly more accented ( $F[2129] = 5.16, p = .023$ ) in the complex than the simple task. Despite the relatively small *task* differences, increasing the demands of the task seemed to detrimentally affect their pronunciation globally. *Task sequence* did not appear to have a significant main effect on comprehensibility ( $F[2129] = 1.22, p = .269$ ) or accentedness ( $F[2129] = 3.62, p = .057$ ) (see Appendix F).

*Associations between acoustic measures and global ratings*

The relationship between acoustic measures (i.e., VOT, vowel quality) and pronunciation ratings (i.e., comprehensibility, accentedness) was assessed through Spearman

**Table 5.** Comprehensibility and accentedness ratings by task.

	Simple				Complex			
	<i>M</i>	<i>SD</i>	95% CI		<i>M</i>	<i>SD</i>	95% CI	
			Lower	Upper			Lower	Upper
Comprehensibility	6.91	0.93	6.70	7.11	6.83	0.83	6.64	7.00
Accentedness	5.08	1.29	4.80	5.36	5.18	1.24	4.90	5.45



**Table 6.** Spearman-rank order correlations between acoustic measures and global ratings by task.

		VOT	Vowel contrastiveness	Vowel nativelikeness
Simple	Comprehensibility	.34*	.24*	-.23*
	Accentedness	-.50**	-.29*	.20
Complex	Comprehensibility	.44**	.37**	-.25*
	Accentedness	-.55**	-.33**	.11

Note:

\* $p < .05$  (2-tailed);

\*\* $p < .001$  (2-tailed).

rank-order correlation coefficients (for subjects with valid data for all measures,  $N = 77$ )<sup>4</sup>. VOT was moderately related to comprehensibility ( $r[154] = .37, p < .001$ ) and accentedness ( $r[154] = -.51, p < .001$ ) ratings, indicating that learners with longer VOT were perceived to be more comprehensible and less strongly accented. In terms of vowel quality, comprehensibility was weakly associated with Mahalanobis distances between contrastive vowels ( $r[154] = .30, p < .001$ ) and with respect to NS ( $r[154] = -.25, p = .002$ ), suggesting that learners producing a larger contrast between the target vowels and producing them more accurately were perceived to be more comprehensible, whereas accentedness was weakly associated with Mahalanobis distances between contrastive vowels only ( $r[154] = -.30, p < .001$ ). These associations were stronger in the complex than the simple task (see Table 6), suggesting that increased task demands strengthened these relationships, especially between VOT and accentedness.

## Discussion

The current study did not find significant effects of task complexity on speech production in terms of lexis, grammar, and fluency (except for the duration of pauses at AS unit boundaries), against the predictions of the cognition hypothesis (Robinson, 2001a, 2011) and the outcomes of previous studies (e.g., Révész, 2009). The effects were in the expected direction, but very small, which may be attributed to the nature of the speaking task, specifically designed to elicit target L2 sounds for acoustic measurement. This design provided learners with plenty of linguistic resources to perform the task, which most likely washed out task complexity effects that could have otherwise emerged. In addition, the rather advanced proficiency of L2 learners might have minimized the internal competition for attentional resources in the areas of complexity, accuracy, and fluency (Skehan, 2015) during L2 oral performance. Investigating L2 speaking fluency, accuracy, and complexity from a dynamic approach (e.g., De Jong, 2023) could also provide further evidence of trade-off effects (e.g., fluency being impeded by difficulty in complex word retrieval) operating during oral production on a very small timescale.

In the present study, detrimental task complexity effects on learners' pronunciation (RQ1: consonant production, RQ2: vowel production RQ3: global pronunciation

<sup>4</sup>The data from the 77 subjects included in the correlation analyses contained all target words. This decision was motivated by the fact that NL' evaluation of comprehensibility and accentedness may have been based greatly in learners' productions of the proper names, which were the most frequent and the ones that differed between simple and complex tasks. Unselecting these meant a loss of 17 participants and affected NL' ratings significantly.

ratings) were predicted, as complexifying tasks along resource-directing dimensions in pronunciation-unfocused tasks could draw learners' attention away from phonological form because high task demands could pose serious limitations on learners' ability to efficiently monitor their speech (Kormos, 1999, 2000). We thus expected Robinson's (2001a, 2011) cognition hypothesis not to hold for pronunciation accuracy (Kuiken & Vedder, 2011; Mora *et al.*, 2024). Such expectations were partly met, as learners' VOT were more target-like in the simple than the complex task (RQ1), suggesting that increased task complexity during authentic communication interfered with learners' ability to focus on segmental features of speech (Derwing *et al.*, 1998). In other words, without an explicit focus on phonetic form, task complexity may not help enhance L2 pronunciation accuracy (e.g., Crowther *et al.*, 2018). These findings align with Skehan's (2009, 2015) and Kormos' (1999, 2000) theory of attentional resource competition during L2 speech performance. Due to limited attentional capacity, learners' attentional resources might have been divided between lexis, grammatical, and pronunciation accuracy, coming into competition during task performance. Although this study cannot provide solid evidence for a lexico-grammatical-pronunciation accuracy trade-off, previous studies have demonstrated trade-offs within one dimension (e.g., syntactic complexity). For example, Wang and Skehan (2014) and Skehan and Shum (2014) showed that subordination and phrasal complexity did not overlap when assessing the role of task structure. Because of the interdependence between linguistic areas, further studies should continue to investigate whether increasing task complexity might have a differential impact on different types of accuracy (i.e., lexical, grammatical, pragmatic, and pronunciation) while learners perform a communicative oral task.

For vowel production, no task complexity effects were found (RQ2). However, given the nature of the speech samples on which vowel quality was measured, and the fact that each subject contributed with a differing number of vowel tokens and words in the simple and complex tasks, no clear interpretation can be drawn from the current findings. Therefore, task complexity effects observed in both learners and NS affecting /æ/ and /ʌ/ could in fact be a consequence of the unbalanced distribution of vowel tokens across words generated by the spontaneity of the task. In addition, the malleability of vowel articulation in terms of the contrastiveness and nativelikeness measures we used, especially in advanced learners, may be too limited to be able to capture any differences resulting from task complexity manipulation in pronunciation-unfocused tasks. VOT, on the other hand, may be a more malleable and salient feature and more readily affected by attentional resources being available to focus on pronunciation.

Task complexity effects on comprehensibility and accentedness (RQ3) were small, although they pointed in the direction of these global dimensions being negatively affected by task complexity (in line with what other studies had found, e.g., Crowther *et al.*, 2018). Interestingly, our analysis revealed differences in learners' comprehensibility and accentedness attributable to task complexity despite there not being significant effects in terms of lexis, grammar, and fluency. The fact that increased task demands did not significantly affect learners' comprehensibility can be partly related to the nature of comprehensibility as a multidimensional construct, including segmentals and suprasegmentals, fluency, lexical and grammatical accuracy and richness (e.g., Trofimovich & Isaacs, 2016).

Regarding RQ4, acoustic and global measures of pronunciation accuracy were expected to be related in both tasks, especially in complex tasks, where task complexity may pose greater demands on L2 pronunciation. Results revealed a weak-to-moderate relation between segmental accuracy, comprehensibility, and accentedness. Learners who produced greater VOT in initial oral stops were perceived to be more

comprehensible and less foreign-accented (i.e., more native-like), in agreement with Riney and Takagi's (1999) findings. Regarding vowel quality, larger Mahalanobis distances between contrastive vowels appeared to be related to higher comprehensibility and lower accentedness ratings, suggesting that the more distinctly learners produced the target vowel contrasts /i:/-/ɪ/ and /æ/-/ʌ/, the more comprehensible and less accented their speech was judged to be, providing some support for a relationship between acoustic distances of contrastiveness in vowel production and NL' ratings of comprehensibility and accentedness. Previous research has found for measures of formant frequencies (Chan et al., 2016; Munro, 1993; Porretta, Kyröläinen & Tucker, 2015). Such associations were found to be stronger in the complex than the simple task (see Crowther et al., 2018), especially for accentedness and VOT.

In sum, the results of the current study revealed that increased task complexity interfered with learners' ability to focus on laryngeal timing (VOT) as a segmental feature of speech, leading to less-target like /p, t, k/ productions in the complex task. However, no significant effects of task complexity were found for vowel production. The findings also indicated small effects of task complexity on comprehensibility and accentedness, suggesting a potential negative effect on these global dimensions. Furthermore, a link was observed between segmental accuracy and the global dimensions of L2 speech, with greater VOT associated with higher comprehensibility and lower accentedness. The associations between acoustic and listener-based measures were stronger in the complex task.

Several methodological limitations suggest future research directions. For example, we were unable to control for the distribution of vowel tokens and words across tasks due to differences in word frequency across tasks and participants. The study also presented variability in the type of phonetic contexts in which target vowels were produced and measured, due to the nature of the oral production task. Whether and how this might have affected the acoustic measures and the comparison between simple and complex tasks is an empirical question warranting future research. It is possible that a more controlled communicative task (e.g., a task imposing a limitation on the use or repetition of vowel tokens) could have yielded more consistent vowel production and a more reliable interpretation of the results. Future studies should try to use more controlled oral communicative tasks involving greater control over word frequency and use of cognates, and more control over the nature of the tokens and the phonetic contexts (e.g., a task requiring the use of adjectives with similar consonantal contexts). More broadly, more experimental research is needed to gain a better understanding of how the manipulation of other task features (i.e., repetition, modality) may affect L2 pronunciation in unfocused tasks. A follow-up of the current study is to analyze potential trade-offs between lexico-grammatical accuracy and pronunciation during monologic/dialogic unfocused oral tasks (see Mora et al., 2024). Finally, in order to ensure the generalizability of the present findings, the study should be replicated with other groups of L2 learners with different ages, L1 backgrounds, proficiency levels, and experience. Controlling for individual differences in aptitude and speaking anxiety may also contribute to our understanding of the relation between task complexity and pronunciation.

## **Conclusion**

The present study set out to explore the relationship between task complexity, pronunciation accuracy (vowel and consonant productions), and listener-based assessments of

L2 speech (comprehensibility and accentedness) in pronunciation-unfocused tasks. The results suggest that complex tasks appeared to hinder the accurate production of English oral stops (e.g., VOT was less accurate in the complex than in the simple task) as well as comprehensibility and accentedness (e.g., speech was rated as less comprehensible and more accented by NL in the complex task), but no observable negative effects of task complexity were observed for vowel production.

In terms of pedagogical implications, future spontaneous tasks should be thoughtfully designed and manipulated to raise learners' awareness about difficult L2 pronunciation targets while communicating. Increasing cognitive complexity in tasks whose pronunciation targets are essential for task completion (e.g., Mora-Plaza, 2023) is one way to promote a focus on phonetic form during interaction. Task-based pronunciation teaching holds a promising avenue for enhancing L2 pronunciation learning in communicative EFL classrooms.

**Data availability statement.** The experiment in this article earned Open Materials badges for transparent practices. The materials and data are available at <https://www.sla-speech-tools.com/>

**Acknowledgments.** The authors would like to thank Athenea Botey and Gonzalo Bermejo (University of Barcelona) for data transcription and annotation, Dr. Danielle Daidone (University of North Carolina Wilmington) for sharing and helping with Praat scripts to automatize acoustic analyses, Dr Joan Borràs-Comes (Laboratori de Fonètica “Eugenio Martínez Celdrán” at the University of Barcelona) for help with data visualization, and Dr Roger Gilabert (University of Barcelona) for expert advice on task complexity. The authors are also thankful to the project members Miren Adrian, Josh Frank, Natalia Fullana, Valeria Galimberti, and Gisela Sosa for their assistance in the study design, data collection, and analysis, and the three anonymous reviewers and handling editor for their insightful comments and helpful suggestions on earlier versions of the manuscript.

**Funding statement.** This study was supported by grants PID2019-107814GB-I00 and PID2022-138129NB-I00 from the Spanish Ministry of Science, Innovation and Universities, and by grant 2023SGR00303 from the Catalan Agency for Management of University and Research Grants (AGAUR).

**Competing interest.** The authors declare that they have no known competing financial, professional, contractual interests or personal relationships that could appear to influence the work reported in this paper.

## References

- Baralt, M., Gurzynski-Weiss, L., & Kim, Y. (2016). Engagement with the language: How examining learners' affective and social engagement explains successful learner-generated attention to form. In M. Sato & S. Ballinger (Eds.), *Peer interaction and L2 learning* (pp. 209–239). Amsterdam, Netherlands: John Benjamins. <https://doi.org/10.1075/llt.45.09bar>
- Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (Version 6.2.19) [Computer program]. Retrieved from <http://www.praat.org/>
- Castañeda, M. L. (1986). El VOT de las oclusivas sordas y sonoras españolas. *Estudios de Fonética Experimental*, 2, 91–110.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372–387. <https://doi.org/10.1016/j.wocn.2005.08.003>
- Cebrian, J. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America*, 145(1), EL52–EL58. <https://doi.org/10.1121/1.5087645>
- Cebrian, J., Gorba, C., & Gavalda, N. (2021). When the easy becomes difficult: Factors affecting the acquisition of the English /i:/-/ɪ/ contrast. *Frontiers in Communication*, 6, 660917. <https://doi.org/10.3389/fcomm.2021.660917>
- Chan, K. Y., Hall, M. D., & Assgari, A. A. (2016). The role of vowel formant frequencies and duration in the perception of foreign accent. *Journal of Cognitive Psychology*, 29(1), 23–34. <https://doi.org/10.1080/20445911.2016.1170746>

- Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, 40(2), 443–457. <https://doi.org/10.1017/S027226311700016X>
- Darcy, I., Mora, J. C. & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, 66 (4), 741–773. <https://doi.org/10.1111/lang.12161>
- De Jong, N. H. (2023). Fluency in Speaking as a Dynamic Construct. *Language Teaching Research Quarterly*, 37, 179–187. <https://doi.org/10.32038/ltreq.2023.37.09>
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410. <https://doi.org/10.1111/0023-8333.00047>
- Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, 27(1), 47–55. <https://doi.org/10.1017/S0025100300005417>
- Docherty, G. J. (1990). An experimental phonetic study of the timing of voicing in English obstruents. *Dissertation Abstracts International*, 51(3).
- Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221–246. <https://doi.org/10.1111/j.1473-4192.2009.00231.x>
- Fllege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.
- Gilabert, R. (2007). The simultaneous manipulation of task complexity along (+/- planning time and here-and-now): Effects on L2 oral production. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44–68). Clevedon, UK: Multilingual Matters. <https://doi.org/10.21832/9781853599286-006>
- Gilabert, R., Barón, J., & Llanes, Á. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, 47, 367–395. <https://doi.org/10.1515/iral.2009.016>
- Corba, C., & Cebrian, J. (2021). The role of L2 experience in L1 and L2 perception and production stops by English learners of Spanish. *Journal of Phonetics*, 88, 101094. <https://doi.org/10.1016/j.wocn.2021.101094>
- Gordon, J. (2021). Pronunciation and task-based instruction: Effects of a classroom intervention. *RELC Journal*, 52(1), 94–109. <https://doi.org/10.1177/00336882209869>
- Grurzynski-Weiss, L., Long, A. Y., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition*, 39(2), 213–224. <https://doi.org/10.1017/S0272263117000080>
- Ishikawa, T. (2008). The effect of task demands of intentional reasoning on L2 speech performance. *The Journal of Asia TEFL*, 5(1), 29–63.
- Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, 63(2), 330–367. <https://doi.org/10.1111/lang.12008>
- Kahneman, D. (1973). *Attention and Effort*. Upper Saddle River, NJ: Prentice Hall.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832. <http://doi.org/10.1121/1.4926561>
- Kim, Y., & McDonough, K. (2008). The effect of interlocutor proficiency on the collaborative dialogue between Korean as a second language learners. *Language teaching research*, 12(2), 211–234. <https://doi.org/10.1177/1362168807086288>
- Kormos, J. (1999). Monitoring and self-repair in L2. *Language learning*, 49(2), 303–342. <https://doi.org/10.1111/0023-8333.00090>
- Kormos, J. (2000). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition*, 22(2), 145–167. <https://doi.org/10.1017/S0272263100002011>
- Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 91–104). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/tblt.2.09ch4>
- Long, M. H. (2015). *Second language acquisition and task-based language teaching*. West Sussex, UK: Wiley Blackwell.
- Márquez, D., & Barón, J. (2021). Effects of task complexity on L2 suggestions: An exploratory study on trade-offs between accuracy and complexity. *TASK*, 1(2), 227–265. <https://doi.org/10.1075/task.20007.mar>

- Meara, P., & Miralpeix, I. (2015). V\_YesNo lognostics vocabulary test. [http://www.lognostics.co.uk/tools/V\\_YesNo/V\\_YesNo.htm](http://www.lognostics.co.uk/tools/V_YesNo/V_YesNo.htm)
- Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2022). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/–/y/ contrast. *Second Language Research*, 38(3), 581–605. <https://doi.org/10.1177/0267658320988061>
- Michel, M., Révész, A., Shi, D., & Li, Y. (2019). The effects of task demands on linguistic complexity and accuracy across task types and L1/L2 speakers. In E. Zhen & M. Ahmadian (Eds.), *Researching L2 task performance and pedagogy: In honour of Peter Skehan* (pp. 133–151). John Benjamins. <https://doi.org/10.1075/tblt.13.07mic>
- Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau, & A. Bel, (eds.), *A portrait of the young in the new multilingual Spain* (pp. 241–263). Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781847690241-015>
- Mora, J. C., & Fullana, N. (2007, August). Production and perception of English /i:/-/ɪ/ and /æ/-/ʌ/ in a formal setting: Investigating the effects of experience and starting age. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1613–1616). Saarbrücken: Universität des Saarlandes.
- Mora, J. C., & Mora-Plaza, I. (2019) Contributions of cognitive attention control to L2 speech learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen & M. H. Sørensen (eds.) *A sound approach to language matters. In honor of Ocke-Schwen Bohn* (pp. 477–499). Dept. of English, School of Communication & Culture, Aarhus University, Denmark. <https://doi.org/10.7146/aul.322.218>
- Mora, J. C., Mora-Plaza, I., & Bermejo Miranda, G. (2024). Speaking Anxiety and Task Complexity Effects on Second Language Speech. *International Journal of Applied Linguistics*, 34(1), 292–315. <https://doi.org/10.1111/ijal.12494>
- Mora, J. C., Rochdi, Y., & Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness*, 23(1–2), 57–75. <https://doi.org/10.1080/09658416.2013.863898>
- Mora-Plaza, I. (2023). Task-based pronunciation teaching helps to improve L2 vowel production: Generalisation effects. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 172–187). Université Grenoble-Alpes. <https://doi.org/10.5281/zenodo.8225325>
- Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). Tools for second language speech research and teaching. <http://sla-speech-tools.com>. <http://doi.org/10.17616/R31NJNAX>
- Mora-Plaza, I., Mora, J. C., & Gilbert, R. (2018). Learning L2 pronunciation through communicative tasks. In J. Levis, C. Nagle & E. Today (Eds.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching Conference* (pp. 174–184). Ames, IA: Iowa State University.
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and speech*, 36(1), 39–66. <https://doi.org/10.1177/002383099303600103>
- Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41, 407–420. <https://doi.org/10.1016/j.wocn.2013.07.005>
- Olson, D. J. (2020). Short-term sources of cross-linguistic phonetic influence: Examining the role of linguistic environment. *Languages*, 5(4), 43. <https://doi.org/10.3390/languages5040043>
- Porretta, V., Kyröläinen, A. J., & Tucker, B. V. (2015). Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics*, 77, 2438–2451. <https://doi.org/10.3758/s13414-015-0916-3>
- Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, 40(3), 491–508. <https://doi.org/10.1016/j.wocn.2012.01.001>
- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, 31(3), 437–470. <https://doi.org/10.1017/S0272263109090366>
- Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, 49(2), 275–302. <https://doi.org/10.1111/0023-8333.00089>
- Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.012>
- Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. <https://doi.org/10.1093/applin/22.1.27>
- Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis*




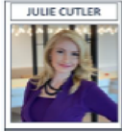






- of language learning and performance (pp. 3–37). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/tblt.2.05ch1>
- Robinson, P., & Gilabert, R. (Eds.). (2007). Task complexity, the Cognition Hypothesis and second language instruction [Special issue]. *International Review of Applied Linguistics*, 45(3). <https://doi.org/10.1515/IRAL.2007.007>
- Sample, E., & Michel, M. (2015). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, 31(8), 23–46. <https://doi.org/10.18806/tesl.v31i0.1185>
- Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In J. J. Ohala, Y. Hasegawa, M. Ohal, D. Granville & A. C. Bailey (Eds.), *Proceedings of the 14<sup>th</sup> International Conference on phonetic sciences* (pp. 607–610). San Francisco, CA: University of California.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 123–156). Amsterdam: John Benjamins. <https://doi.org/10.1075/tblt.8.05ske>
- Skehan, P., & Shum, S. (2014). Structure and processing condition in video-based narrative retelling. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 187–210). Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tblt.5.07ske>
- Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, 39(2), 347–380. <https://doi.org/10.1017/S0272263116000425>
- Sudharshana, N. P. (2021). From cognitive grammar to pedagogic grammar: Macrostrategies for designing form-focused tasks. In N. P. Sudharshana, & L. Mukhopadhyay (Eds.), *Task-Based Language Teaching and Assessment: Contemporary Reflections from Across the World* (pp. 163–181). Singapore: Springer. [https://doi.org/10.1007/978-981-16-4226-5\\_9](https://doi.org/10.1007/978-981-16-4226-5_9)
- Trofimovich, P., & Isaacs, T. (2016). Second language pronunciation assessment: A look at the present and the future. In T. Isaacs & P. Trofimovich (Eds.), *Second Language Pronunciation Assessment: Interdisciplinary Perspectives* (pp. 259–271). Multilingual Matters. <https://doi.org/10.21832/ISAACS6848>
- Ur, P. (1981). *Discussions that work: Task-centered fluency practice*. Cambridge, UK: Cambridge University Press.
- Van den Branden, K. (Ed.) (2006). *Task-based language education: From theory to practice*. Cambridge, UK: Cambridge University Press.
- Wang, Z., & Skehan, P. (2014). Structure, lexis, and time perspective: Influences on task performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 155–185). Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tblt.5.06wan>
- Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Human skills* (pp. 71–105). John Wiley & Sons.
- Willis, J. (1996). *A Framework for Task-Based Learning*. Harlow: Longman.
- Wu, S. L., Tio, W. P., & Ortega, L. (2021). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies of Second Language Acquisition*, 44(1), 271–300. <https://doi.org/10.1017/S0272263121000103>

**Appendices**

**Appendix A. Pre-task: Instructions and materials**

Listen carefully to the description of the following characters and link them to the words that describe their personalities and professions by putting their numbers next to the words.

Character																																																																												
1				<table border="1" style="border-collapse: collapse; width: 100%;"> <tr><td> </td><td>actress</td></tr> <tr><td> </td><td>banker</td></tr> <tr><td> </td><td>bitchy</td></tr> <tr><td> </td><td>catholic</td></tr> <tr><td> </td><td>charismatic</td></tr> <tr><td> </td><td>deceitful</td></tr> <tr><td> </td><td>determined to win</td></tr> <tr><td> </td><td>dreamy</td></tr> <tr><td> </td><td>dull</td></tr> <tr><td> </td><td>funny when tipsy</td></tr> <tr><td> </td><td>genius</td></tr> <tr><td> </td><td>greedy</td></tr> <tr><td> </td><td>humble</td></tr> <tr><td> </td><td>impulsive</td></tr> <tr><td> </td><td>judge</td></tr> <tr><td> </td><td>keen on martial arts</td></tr> <tr><td> </td><td>kickboxing</td></tr> <tr><td> </td><td>kung fu</td></tr> <tr><td> </td><td>liberal</td></tr> <tr><td> </td><td>mean</td></tr> <tr><td> </td><td>painting</td></tr> <tr><td> </td><td>passionate</td></tr> <tr><td> </td><td>peaceful</td></tr> <tr><td> </td><td>poker</td></tr> <tr><td> </td><td>politician</td></tr> <tr><td> </td><td>politician</td></tr> <tr><td> </td><td>popular</td></tr> <tr><td> </td><td>pragmatic</td></tr> <tr><td> </td><td>silly</td></tr> <tr><td> </td><td>simple and kind</td></tr> <tr><td> </td><td>stubborn</td></tr> <tr><td> </td><td>table tennis</td></tr> <tr><td> </td><td>talented artist</td></tr> <tr><td> </td><td>teacher</td></tr> <tr><td> </td><td>trustworthy</td></tr> <tr><td> </td><td>wine-tasting</td></tr> </table>		actress		banker		bitchy		catholic		charismatic		deceitful		determined to win		dreamy		dull		funny when tipsy		genius		greedy		humble		impulsive		judge		keen on martial arts		kickboxing		kung fu		liberal		mean		painting		passionate		peaceful		poker		politician		politician		popular		pragmatic		silly		simple and kind		stubborn		table tennis		talented artist		teacher		trustworthy		wine-tasting
	actress																																																																											
	banker																																																																											
	bitchy																																																																											
	catholic																																																																											
	charismatic																																																																											
	deceitful																																																																											
	determined to win																																																																											
	dreamy																																																																											
	dull																																																																											
	funny when tipsy																																																																											
	genius																																																																											
	greedy																																																																											
	humble																																																																											
	impulsive																																																																											
	judge																																																																											
	keen on martial arts																																																																											
	kickboxing																																																																											
	kung fu																																																																											
	liberal																																																																											
	mean																																																																											
	painting																																																																											
	passionate																																																																											
	peaceful																																																																											
	poker																																																																											
	politician																																																																											
	politician																																																																											
	popular																																																																											
	pragmatic																																																																											
	silly																																																																											
	simple and kind																																																																											
	stubborn																																																																											
	table tennis																																																																											
	talented artist																																																																											
	teacher																																																																											
	trustworthy																																																																											
	wine-tasting																																																																											
2																																																																												
3																																																																												
4																																																																												
5																																																																												
6																																																																												

**Appendix B. The dinner table task: Instructions and materials**

In this speaking task, we ask you to organize a successful dinner party for six people.

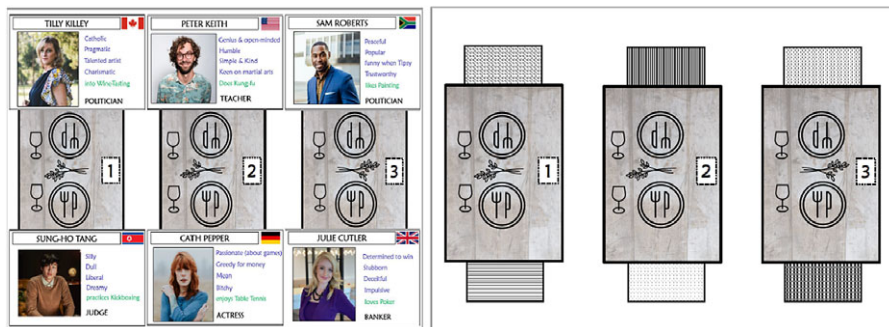
Please read the information about each person that is coming to the party carefully.

Your goal is:

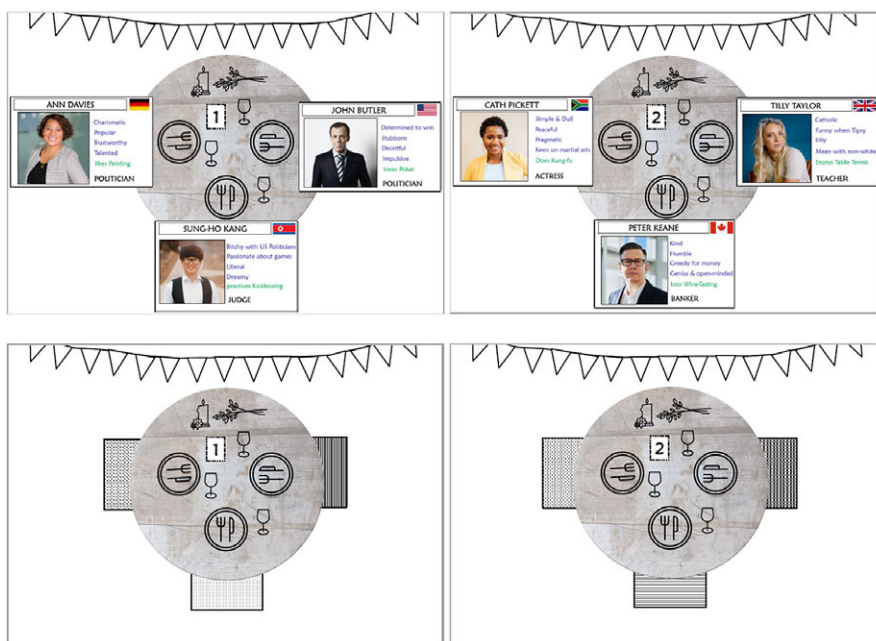
1. To justify why the following seating arrangement will not guarantee a successful dinner party.
2. To provide and justify a new seating arrangement that can guarantee a successful dinner party: (1) take the individual cards, (2) place them on the chairs of the new table.

Refer to characters with names and surnames.

Foster smooth and pleasant conversations between the guests.



Simple task: part 1 & part 2



Complex task: part 1 and part 2

\*The printable materials can be found online: <http://sla-speech-tools.com>

**Appendix C. Task performance questionnaire**

Q1. How **well** do you think you have performed the task?

1 2 3 4 5 6 7 8 9  
Very poorly Very well

Q2. How **difficult** did you find the task?

1 2 3 4 5 6 7 8 9  
Not difficult at all Extremely difficult

Q3. Did you feel **anxious** when performing the task?

1 2 3 4 5 6 7 8 9  
I did not feel anxious at all I felt very anxious

Q4. How much **mental effort** did you have to put to perform this task?

1 2 3 4 5 6 7 8 9  
No mental effort at all Extreme mental effort

**Appendix D. List of target words by task, vowel, and consonant**

	Vowels				Consonants		
	/æ/	/ʌ/	/ɪ/	/i:/	/p/	/t/	/k/
Simple and complex	actress	dull	bitchy	deceitful	painting	determined	kung fu
	banker	funny	kickboxing	dreamy	poker	table tennis	catholic
	catholic	humble	liberal	genius	politician	wine-tasting	keen
	charismatic	impulsive	silly	greedy	popular	talented	kickboxing
	passionate	judge	simple	keen	passionate	teacher	kind
	pragmatic	stubborn	politician	mean	peaceful	tipsy	Cath
	talented	trustworthy	Tilly	peaceful	impulsive	Tilly	
	Cath	Sung-ho	tipsy	teacher	Peter		
Simple	Sam	Cutler	Killey	Keith	Pepper	Tang	Cutler
	Tang						Killey
Complex	Ann	Butler	Pickett	Keane	Pickett	Taylor	Keith
	Kang						Kang
							Keane



**Learners' oral stops (voice onset time)**

VOT	$\beta$	SE	<i>t</i>	<i>p</i>	95% CI	
					Lower	Upper
<i>Intercept</i>	3.703	0.039	95.309	0.000	3.627	3.779
<i>Task</i>	0.061	0.015	4.040	0.000	0.032	0.091
<i>Consonant (/p/)</i>	-0.614	0.015	-39.629	0.000	-0.644	-0.583
<i>Consonant (/t/)</i>	0.172	0.015	11.771	0.000	0.143	0.200
<i>Sequence</i>	0.134	0.051	2.651	0.008	0.035	0.233
<i>Task [S]* Consonant (/p/)</i>	-0.025	0.022	-1.122	0.262	-0.068	0.018
<i>Task [S]* Consonant (/t/)</i>	-0.024	0.021	-1.150	0.250	-0.065	0.017

Reference levels: task = complex; consonant = /k/; sequence = C > S.

**Vowels (Mahalanobis distances)**

Contrastiveness	$\beta$	SE	<i>t</i>	<i>p</i>	95% CI	
					Lower	Upper
<i>Intercept</i>	0.935	0.039	24.268	0.000	0.859	1.011
<i>Task</i>	-0.080	0.028	-2.809	0.005	-0.135	-0.024
<i>Contrast</i>	-0.364	0.025	-14.334	0.000	-0.414	-0.315
<i>Sequence</i>	-0.002	0.047	-0.051	0.959	-0.094	0.090
<i>Task × Contrast</i>	0.053	0.037	1.443	0.149	-0.019	0.125

Reference levels: task = complex; contrast = /æ/-/ʌ/; sequence = C > S.

Nativelikeness	$\beta$	SE	<i>t</i>	<i>p</i>	95% CI	
					Lower	Upper
<i>Intercept</i>	0.395	0.034	11.706	0.000	0.329	0.461
<i>Task</i>	-0.010	0.032	-0.295	0.768	-0.073	0.054
<i>Vowel (/i:/)</i>	-0.272	0.029	-9.262	0.000	-0.330	-0.215
<i>Vowel (/ɪ/)</i>	-0.072	0.029	-2.470	0.014	-0.128	-0.015
<i>Vowel (/æ/)</i>	-0.077	0.032	-2.394	0.017	-0.141	-0.014
<i>Sequence</i>	0.045	0.035	1.267	0.205	-0.025	0.114
<i>Task [S] × Vowel (/i:/)</i>	-0.012	0.042	-0.294	0.768	-0.095	0.070
<i>Task [S] × Vowel (/ɪ/)</i>	0.051	0.042	1.218	0.223	-0.031	0.134
<i>Task [S] × Vowel (/æ/)</i>	-0.050	0.046	-1.094	0.274	-0.140	0.040

Reference levels: task = complex; vowel = /ʌ/; sequence = C > S.



**Comprehensibility**

	$\beta$	SE	t	p	95% CI	
					Lower	Upper
Intercept	6.713	0.1386	48.426	0.000	6.441	6.985
Task	0.081	0.0418	1.930	0.054	-0.001	0.163
Sequence	0.204	0.1850	1.105	0.269	-0.158	0.567

Reference levels: task = complex; sequence = C > S.

**Accentedness**

	$\beta$	SE	t	p	95% CI	
					Lower	Upper
Intercept	5.464	0.2023	27.012	0.000	5.067	5.860
Task	-0.096	0.0421	-2.273	0.023	-0.178	-0.013
Sequence	-0.517	0.2716	-1.903	0.057	-1.049	0.016

Reference levels: task = complex; sequence = C > S.

**Cite this article:** Mora-Plaza, I., Mora, J. C., Ortega, M., & Aliaga-Garcia, C. (2024). Is L2 pronunciation affected by increased task complexity in pronunciation-unfocused speaking tasks? *Studies in Second Language Acquisition*, 1–33. <https://doi.org/10.1017/S0272263124000470>