Mora, J. C. & Levkina, M, (2018). Training vowel perception through map tasks: The role of linguistic and cognitive complexity. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, University of Utah, September, 2017 (pp. 151-162). Ames, IA: Iowa State University.

TRAINING VOWEL PERCEPTION THROUGH MAP TASKS: THE ROLE OF LINGUISTIC AND COGNITIVE COMPLEXITY.

Joan C. Mora, University of Barcelona

Mayya Levkina, University of Barcelona

Recent research suggests that manipulating task design variables promotes attention to phonetic form during communicative interaction, leading to accuracy gains in production. This study investigated the effect of linguistic and cognitive complexity sequencing on the perception of a difficult L2 vowel contrast for L1-Spanish learners of English (/iː/-/ɪ/). The L2-English learners (*n*=81) were randomly assigned either to an experimental training group (EG, *n*=66) or a control group (CG, *n*=15). EG participants performed computerized map tasks that required them to give and follow directions to map locations that could only be successfully reached by accurately perceiving and producing the intended target contrast. Accuracy in perception was assessed before and after training through a categorical ABX discrimination task. The results revealed gains in discrimination accuracy and speed for EG participants, and treatment and experimental condition effects. Overall the results underscore the potential of task-based phonetic form-focused instruction for L2 pronunciation development.

INTRODUCTION

Most second language (L2) learners find pronunciation difficult to master and tend to exhibit strong L1 accents when speaking the L2, potentially causing comprehensibility problems when interacting with native speakers (Moyer, 1999). Similarly, L2 pronunciation remains a challenge for L2 teachers, who find pronunciation teaching difficult to integrate into their communication-oriented L2 classes (Darcy, Ewert, & Lidster, 2012). Still, empirical research on the effectiveness of task design and manipulation in L2 pronunciation teaching remains scarce (Gurzynski-Weiss, Long, & Solon, 2017), evidence of the difficulty in bridging the gap between research on L2 phonology and pronunciation teaching (Darcy, 2017). One of the difficulties of integrating pronunciation teaching into a communicative language teaching framework lies in designing pedagogic tasks that make a focus on phonetic form during student peer interaction essential for task completion (Mora & Levkina, 2017). Research has demonstrated that a focus on phonetic form and explicit corrective feedback may facilitate learners' awareness of L2 sound structure, possibly leading to benefits in L2 pronunciation development (Saito & Wu, 2014). One approach that is promising in bridging the gap between research and practice in L2 pronunciation teaching is task-based pronunciation teaching (TB*P*T), that is, the application of the tenets and principles of task-based language teaching (TBLT) to L2 pronunciation. A TB*P*T approach to pronunciation instruction generates pronunciation-focused communicative tasks by embedding a focus on phonetic form through task design manipulation (Mora & Levkina, 2017). Several recent studies have investigated the potential of TBLT principles and findings to extend beyond grammar and lexis to pronunciation.

In general the findings of these studies suggest that it is possible to manipulate the design parameters of communicative tasks to encourage a focus on phonetic form during communicative interaction, and that this leads to L2 pronunciation development, both at the level of segmental accuracy (Solon, Long, & Gurzynski-Weiss, 2017) and at the suprasegmental level (McKinnon, 2017; Parlak & Ziegler, 2017). For example, Solon et al. (2017) tested the following TBLT predictions about increasing task complexity for L2 pronunciation. First, more complex tasks, as opposed to simpler tasks, should lead to a higher occurrence of pronunciation-based language related episodes (LREs) where students may engage in discussion about the target phonetic forms. Second, according to the *Cognition Hypothesis* (Robinson, 2003) more complex tasks should lead to larger gains in pronunciation accuracy than simpler tasks. Solon et al. set up a map task to be conducted in pairs (17 dyads; *n*=34) by L1-English intermediate level learners of Spanish containing minimal-pair street names targeting the 5 vowel phonemes of Spanish (/i/, /e/, /a/, /o/, /u/). Learners performed two map tasks varying in cognitive complexity, a simple one with few elements and a complex one containing many more elements. Although they found more (though not significantly more) pronunciation-focused LREs in the simple than in the complex task, contrary to their expectations, they found significant improvement in how accurately learners produced Spanish /e/ during the complex task, just as they had predicted.

As previous research within the Cognition Hypothesis (Robinson, 2001) had found for grammar and lexis, Solon et al. (2017) suggests that manipulating task complexity may benefit pronunciation accuracy. However, in their study map tasks were used to induce a focus on phonetic form through the use of minimal-pair street names contrasting vowels that were perceptually distinct (/a, i, e, o, u/) and embedded in a variety of target lexical items. It therefore remains unclear whether task complexity manipulation would have the same effect on confusable contrasting L2 vowels, such as English /iː/-/ɪ/ for L1-Spanish learners, or confusable vowels embedded in target word forms of varying linguistic complexity. In the current study we did not assess inter-group differences in segmental accuracy gains resulting from manipulations of task complexity. Instead, we kept cognitive complexity constant across participant groups by sequencing three map tasks in order of increasing cognitive complexity (simple, + complex, ++ complex), following Robinson's (2010) SSARC model, and manipulated linguistic complexity.

The present study addressed the following three research questions:

RQ1: Will a map task treatment be effective in improving L2 learners' perception of a confusable L2 vowel contrast (/iː/-/ɪ/)?

RQ2: Will linguistic complexity affect the effectiveness of the treatment?

RQ3: Will learners be able to generalize gains to new items and speakers?

THE PRESENT STUDY

According to Robinson's SSARC model, the main criterion in sequencing tasks is the manipulation of cognitive complexity along resource-dispersing and resource-directing variables (Robinson, 2010; Robinson & Gilabert, 2007). In our study, all participants, L1-Spanish learners of English, performed all three map tasks in the same order of cognitive complexity, but they were randomly assigned to one of three groups differing in the type of linguistic complexity of the task. We operationalized lingusitic complexity as the number of syllables a nonword contained. Three-syllable nonwords (*fadeetick-fadittick*) were defined as linguistically more complex than one-syllable nonwords (*deet-ditt*) because they contained more segmental material to process before the target stressed syllable. Perceiving the target vowel contrast /iː/-/ɪ/ in a three-syllable nonword was therefore expected to pose greater difficulty and to require greater attentional effort to learners than perceiving the same contrast in a corresponding 1-syllable nonword, which might lead to increased benefits in developing sensitivity to the contrast. Our aim, therefore, was twofold. First, we assessed the overall effect of a sequence of map tasks on L2 segmental perceptual accuracy. Secondly, we assessed inter-group differences in perceptual accuracy gains as a function of the level of the complexity of the stimuli as defined above. Participants were assigned to one of three groups as a function of the complexity of the nonword items where the target English vowel contrast had been embedded: 1 syllable, 3 syllables, or a mixture of both. We chose the target English vowel contrast /iː/-/ɪ/ (e.g. *beat-bit*) because it has been shown to be difficult for Spanish learners of English to acquire, even at advanced levels of proficiency. The English vowel phonemes /iː/ and /ɪ/ are both perceptually mapped onto a single L1-Spanish phonemic vowel category /i/ (Cebrian, Mora, & Aliaga-Garcia, 2011). It is therefore a case of single category assimilation according to the Perceptual Assimilation Model PAM-L2 (Best & Tyler, 2007), which would predict considerable difficulty in acquisition for this English vowel contrast by L1-Spanish learners.

METHODS

The present study followed a pre-test > treatment > post-test design. All the participants took a perception test (AXB discrimination) and two production tests (delayed nonword and sentence repetition tasks) of the target vowel contrast /iː/-/ɪ/ in trained and untrained items. They also performed a map task test where they were asked to give and follow directions on a map. The participants in the experimental group, unlike those in the control group, did the treatment, administered over two weeks, consisting of perception and production familiarization tasks and three map tasks performed in order of increasing simple-to-complex cognitive complexity. Due to space limitations only the perception data is presented in the present paper.

Participants

The participants were 81 intermediate-to-advanced L1-Spanish learners of English. They had studied English in a foreign language instructional setting in Spain all through primary and secondary education and at the time of testing they were first-year university students starting a degree in philology. Participants were randomly assigned to an experimental group (EG: *n*=66) or to a control group (CG: *n*=15) that did not do any of the treatment tasks. Three participants in the experimental group were excluded from analysis because they missed the post-test. EG participants were further randomly assigned to one of three groups as a function of the type of stimuli they would be exposed to during the training tasks: EG1 (*mixed*) received a balanced mixed exposure to 1- and 3-syllable nonwords, EG2 (*simple*) were only exposed to 1-syllable nonwords and EG3 (*complex*) were only exposed to 3-syllable nonwords.

Stimuli and instruments

Four native speakers of standard Southern British English (2 female) recorded the speech stimuli for all tests and tasks. One female and one male voice were used for the familiarization pre-task and for the map tasks and half of the items in the ABX discrimination test. The other two voices were used for the other half of the items to test for generalization effects. The stimuli, with the target vowels /iː/ and /ɪ/ in stressed position, were either simple one-syllable CVC nonwords (*peef* vs. *piff*), or complex three-syllable CVˈCVCVC nonwords (*lapeefan* vs. *lapiffan*) created by adding phonotactically legal initial CV- and final -VC syllables to monosyllabic CVC nonwords.

ABX discrimination

The testing instrument was an ABX discrimination task. The ABX test contained 16 test nonword triads in 4 orders (ABB, ABA, BAA, BAB), i.e. 64 test trials, and 8 control nonword triads in 4 orders (32 control trials) based on 4 vowel contrasts (/iː/-/æ/, /ɪ/-/æ/, /iː/-/uː/, /ɪ/-/uː/) posing no discrimination difficulty. The test, with a total of 96 trials, was constructed so that in both test and control conditions there was always an equal number of 1- and 3-syllable nonwords, an equal number of trained and untrained nonwords, and an equal number of nonwords spoken by a male and a female voice. In each ABX triad, A and B were spoken by the same voice (either male or female) and X in a different voice (e.g., female if A and B were spoken by a male voice). Participants were instructed to decide whether the last item in a triad (X) was the same as the first (A) or the second (B) item as fast and as accurately as they could. ABX trials were presented with an inter-stimulus interval of 500 ms and an inter-trial interval of 2000 ms upon participant's response or 2500 ms after the onset of the last item in the triad if no response was provided. We obtained individual accuracy and reaction time (RT) scores for each participant across the main trial type conditions: test vs. control; 1 vs. 3 syllables; trained vs. untrained. RTs from wrong responses were excluded from analysis and the remaining RT data were screened by subject (and by condition) for RTs above or below 2.5 standard deviations (SDs) from the subject's mean RT.

Familiarization pre-task

For the familiarization pre-task and the three map tasks L2 learners were randomly assigned to one of three groups as a function of the linguistic complexity of the stimuli they were exposed to. The treatment pre-task consisted of an identification task containing the 48 "trained" nonwords in the ABX task described above. EG1 received exposure to 24 1-syllable (18 test + 6 control) and 24 3-syllable (18 test + 6 control) nonwords (36 test trials), EG2 was exposed to 48 1-syllable nonwords (18 test + 6 control x 2 repetitions; 36 test trials) and EG3 was exposed to 48 3-syllable nonwords (18 test + 6 control x 2 repetitions; 36 test trials). Nonword trials were randomly presented over headphones to participants, who were asked to identify the nonword they had heard by clicking on one of the 24 response buttons labelled in normal orthography. Minimal pairs (e.g. *piff* vs. *peef*) appeared side by side and were alphabetically distributed on the screen. The purpose of this task was to familiarize learners with the auditory and orthographic version of the nonwords they would encounter in the map tasks.

Map tasks

The participants in each of the three experimental groups (EGs) defined by the linguistic complexity of the stimuli (EG1: *mixed*; EG2: *simple*; EG3: *complex*) performed three computerized collaborative map tasks that differed only in the linguistic complexity of the nonwords used as street names (EG1: monosyllabic and trisyllabic; EG2: monosyllabic only; EG3: tri-syllabic only).

The tasks were performed in pairs so that first student A would give directions to student B, and then they would exchange roles. Students were sitting in front of two computer screens, one used for giving directions (1) and one for following directions (2). Student A would give directions to student B (following a marked itinerary on his map on screen 1) to pick a parcel located beyond a crossroad that led to two streets with minimal-pair names (e.g. *lapiffan* / *lapeefan*), each having a grey parcel icon. For example, student A would say: *now please pick the parcel you will find in "lapiffan" street*. Student B would need to decide whether it was the parcel in *lapiffan* street or the one in *lapeefan* street by clicking on it. In order to give participants feedback on accuracy, when clicked, parcels would go green if correct or red if wrong. Student A could always monitor what student B was doing on screen 2, but student B could not see student A's marked itinerary because his/her screen 2 would be switched off.

This situation raised clarification questions and repetition requests involving the contrasting street names, which generated pronunciation-focused LREs. In the maps used in the map tasks, the names of the streets were those used in the familiarization pre-task and were clickable, so that both student A and B could click on them to hear them over headphones. The map task sessions were digitally recorded (44.1kHz, 16-bit) onto Marantz PMD660 recorders separately for each student via Shure SM58 microphones.

RESULTS

The first aim in the current study was to assess perceptual gains in L2 learners' sensitivity to the /iː/-/ɪ/ contrast resulting from the treatment (a sequence of map tasks). We first examined mean differences between test and control trials. At pre-test control trials in the ABX discrimination test obtained significantly higher accuracy scores (proportion of correct responses) than test items did (see Table 1; EGs: $t(6334)=-16.52$, $p<.001$; CG: $t(1438)=-8.97$, $p<.001$) and RTs were significantly faster on control trials than on test trials (EG: $t(4397)=5.97$, $p<.001$; CG: $t(1031)=3.93$, $p<.001$). The same pattern of results was obtained at post-test, suggesting that the test contrast /iː/-/ɪ/ was difficult to discriminate, with mean 0.64-0.70 proportion correct scores. We next examined perceptual gains for the test trials only for all participant groups. In general, results showed differences in the expected direction, with post-test scores reflecting significant improvement in response accuracy ($t(8372)=-5.97$, $p<.001$) and speed ($t(5376)=9.28$, $p<.001$) for the EGs. As expected, improvement in accuracy did not reach significance ($t(1918)=-1.41$, $p=.158$) for the CG,, but response speed did ($t(1261)=4.32$, $p<.001$).

Table 1

Accuracy and RT scores by item type and testing time.

Group	Item Type	Test	Accuracy			RT		
			N	*M*	*SD*	*N*	*M*	*SD*
EGs	Test	Pre	4224	.64	.48	2664	1156.08	363.94
		Post	4160	.70	.46	2858	1069.25	331.33
	Control	Pre	2112	.84	.37	1735	1089.09	362.96
		Post	2080	.90	.30	1831	1001.76	328.46
CG	Test	Pre	960	.66	.47	620	1160.70	332.17
		Post	960	.69	.46	643	1084.15	295.93
	Control	Pre	480	.88	.33	413	1079.50	315.36
		Post	480	.92	.28	431	1013.72	320.34

Secondly, we assessed experimental inter-group differences in perceptual accuracy gains as a function of the level of the complexity of the stimuli for test trials only. Gains in accuracy and speed were larger for experimental subjects in the simple (EG2) and complex (EG3) conditions than for subjects in the mixed condition (EG1), suggesting that exposure to items of varying complexity might have been detrimental to accuracy gains (Figure 1). In order to test this we submitted the accuracy and speed scores to two-way ANOVAs with *Test* (pre-test, post-test) as a within-subjects factor and *Group* (mixed, simple, complex) as the between-subjects factor. The results showed, for accuracy, a significant main effect of *Test*

(F(1,60)=37.99, p<.001, η^2=.388), but neither the main effect of *Group* (F(2, 60)=.141, p=.869, η^2=.005) nor the *Test* x *Group* interaction (F(2, 60)=2.47, p=.092, η^2=.076) reached significance. For speed, the same pattern of results was found, with a significant main effect of *Test* (F(1,60)=28.86, p<.001, η^2=.325) and a non-significant main effect of *Group* (F(2, 60)=2.23, p=.115, η^2=.069). The *Test* x *Group* interaction was not significant (F(2, 60)=.391, p=.678, η^2=.013). These results show that the treatment was effective in improving L2 learners' perception of the L2 vowel contrast /iː/-/ɪ/ (RQ1).

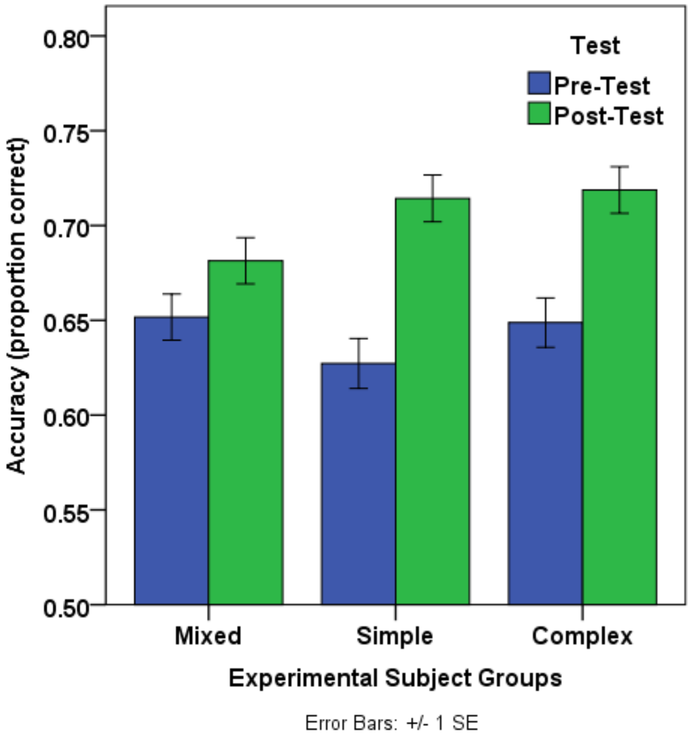


Figure 1. Proportion of correct responses by Experimental Group.

Gain sizes were larger for the groups exposed to the simple and complex treatment types (EG2 and EG3) than for the group exposed to a mixed treatment (EG1), but such group differences based on the linguistic complexity of stimuli did not reach significance (RQ2). This was further confirmed by a series of ANOVAs on gain scores (post-test minus pre-test) for accuracy and response speed (Figure 2). For accuracy gains the main effect of *Group* approached significance (F(2,60)=2.479, p=.092, η^2=.076), and was not significant for response speed (F(2,60)=.391, p=.678, η^2=.013). Accuracy benefits therefore patterned as *Simple* > *Complex* > *Mixed*, whereas RT benefits patterned as *Simple* = *Complex* > *Mixed*, but non-significantly.

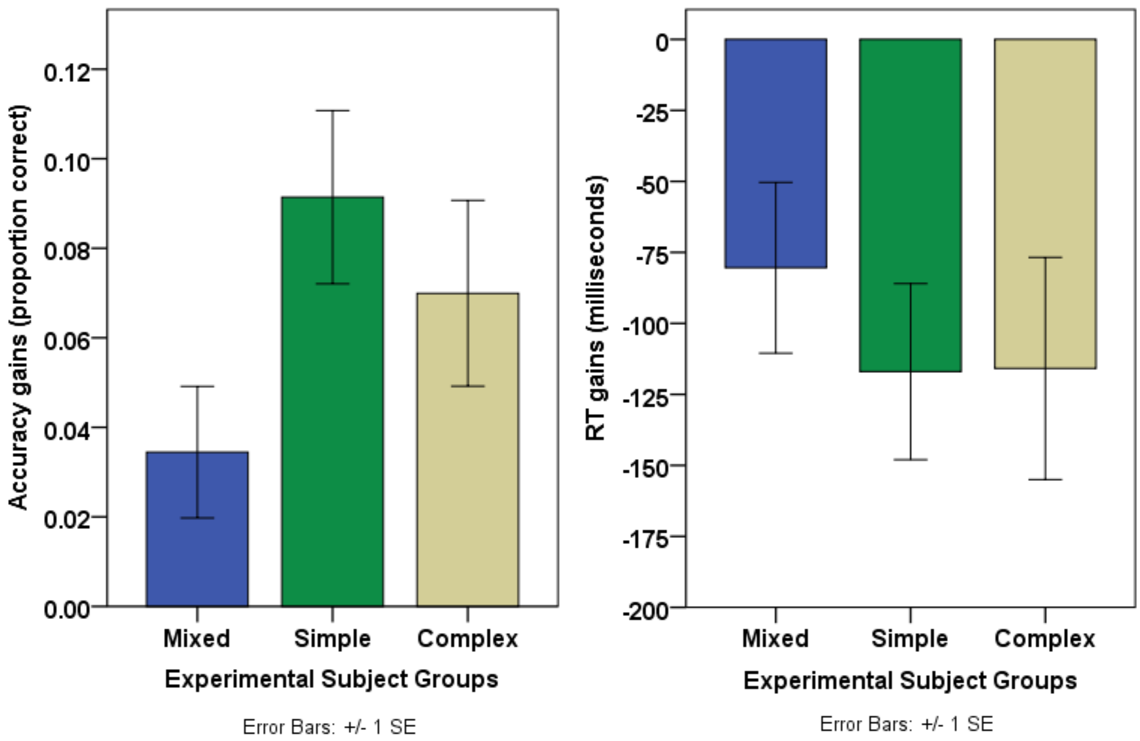


Figure 2. Accuracy and speed gains by experimental group.

Because half of the test items in the ABX discrimination test were monosyllabic (e.g. *piff*) and half were tri-syllabic (e.g. *lapiffan*), and participants had been assigned to different linguistic complexity treatment groups based on whether they were exposed to monosyllabic, tri-syllabic or both types of stimuli, it is likely for the three experimental groups to differ from one another in how accurately and how fast they could correctly identify 1-syllable and 3-syllable test items. In order to explore this possibility we analysed participants' performance separately by test item type (1 or 3 syllables). As Figure 3 shows, EG1 (mixed) obtained the smallest gains in both 1- and 3-syllable nonwords, whereas EG2 (simple) obtained the largest gains on 1-syllable nonwords, which appeared to be as large as those they obtained on 3-syllable nonwords, whereas EG3 (complex) obtained larger gains on 3-syllable words than on 1-syllable words.

We tested these differential gains through two ANOVAs with subject *Group* (Mixed, Simple, Complex) as the between subjects factor and *Test* (pre-test, post-test) and *Trial Type* (1-syllable trials, 3-syllable trials) as within subjects factors, one for response accuracy (proportion correct), and one for response speed (RTs in milliseconds). For both accuracy and speed the main effects of *Group* were not significant ($F(2,60)=.141$, $p=.869$, $\eta^2=.005$; $F(2,60)=2.239$, $p=.115$, $\eta^2=.069$; respectively) whereas the main effects of *Test* ($F(1,60)=37.99$, $p<.001$, $\eta^2=.388$; $F(1,60)=28.86$, $p<.001$, $\eta^2=.325$) and *Trial Type* ($F(1,60)=15.91$, $p<.001$, $\eta^2=.210$; $F(1,60)=246.83$, $p<.001$, $\eta^2=.804$) were significant. None of the interactions reached significance. This suggests that all groups significantly improved from pre- to post-test on both 1- and 3-syllable trials, even if participants generally found 3-syllable trials were harder and slower to discriminate[1]. However, within-group *t*-tests

[1]RTs were measured from the onset of the last nonword in an ABX triad for both 1- and 3-syllable trials. On average RTs on 3-syllable trials were 128 ms (pre-test) and 105 ms (post-test) slower than 1-syllable trials, whereas the initial unstressed syllable in 3-syllable nonwords was approximately 100 ms long. Therefore, response speed differences as a function of linguistic complexity cannot be attibuted to increrased difficulty or complexity, but to the extra response time needed to respond due to the initial syllable before the target contrast.

contrasting 1-syllable vs. 3-syllable mean accuracy and speed scores failed to reveal significance in any of the differences observed. In addition, the size of gains in accuracy (*t*(62)=.066, *p*=.948) and speed (*t*(62)=.1.74, *p*=.0.81) were comparable for 1-syllable and 3-syllable test-trials.

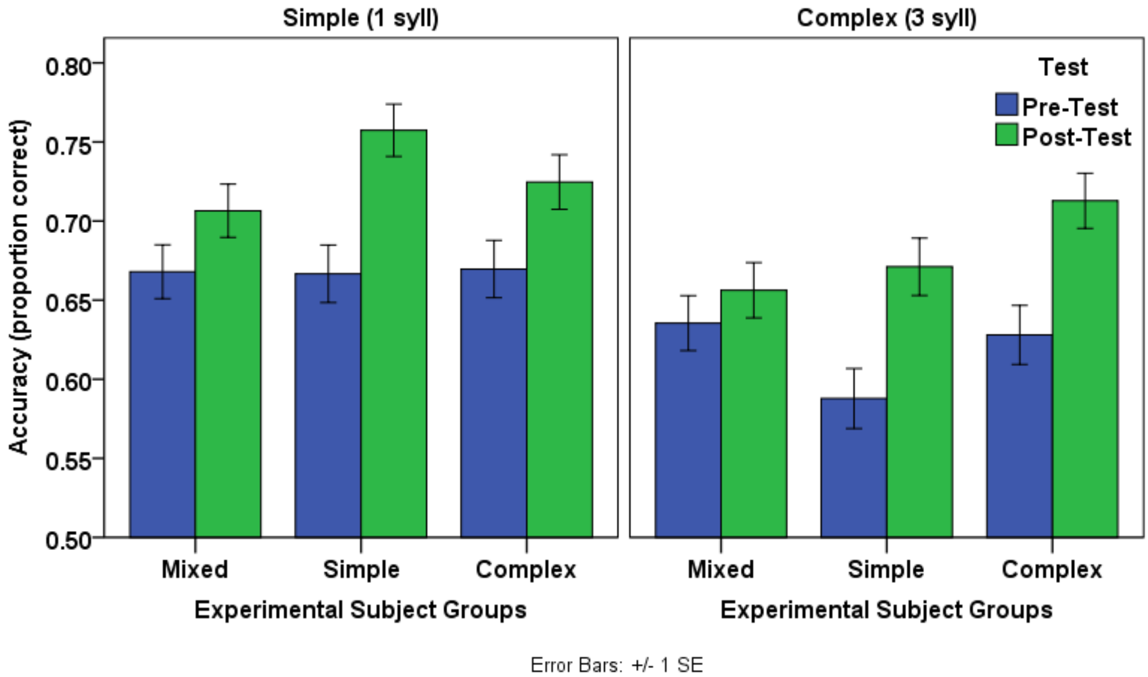


Figure 3. Pre-test and post-test accuracy scores by EG and Trial Type.

Finally, in order to test for generalization effects (RQ3), we assessed the extent to which the general significant gains in accuracy and speed found applied equally to trained and untrained test items. Test items containing untrained nonwords (new nonwords and nonwords spoken in by a new voice) significantly improved in accuracy (*M*=.63>.69; *t*(4187)=-3.86, *p*<.001) and speed (*M*=1158>1067; *t*(2617)=6.69, *p*<.001) from pre-test to post-test. The test items produced in a new voice were also found to improve significantly in accuracy (*M*=.64>.69; *t*(4187)=-3.81, *p*<.001) and speed (*M*=1116>1028; *t*(2670)=6.57, *p*<.001). These results show that participants were able to generalize gains to new contexts ("nonwords") and to new speakers ("voices") suggesting that they had learned to apply the discrimination skills acquired through the training to novel test items. As shown in Figure 4, gains for untrained items were only slightly smaller than those for trained items, as expected, and occurred in all experimental groups.

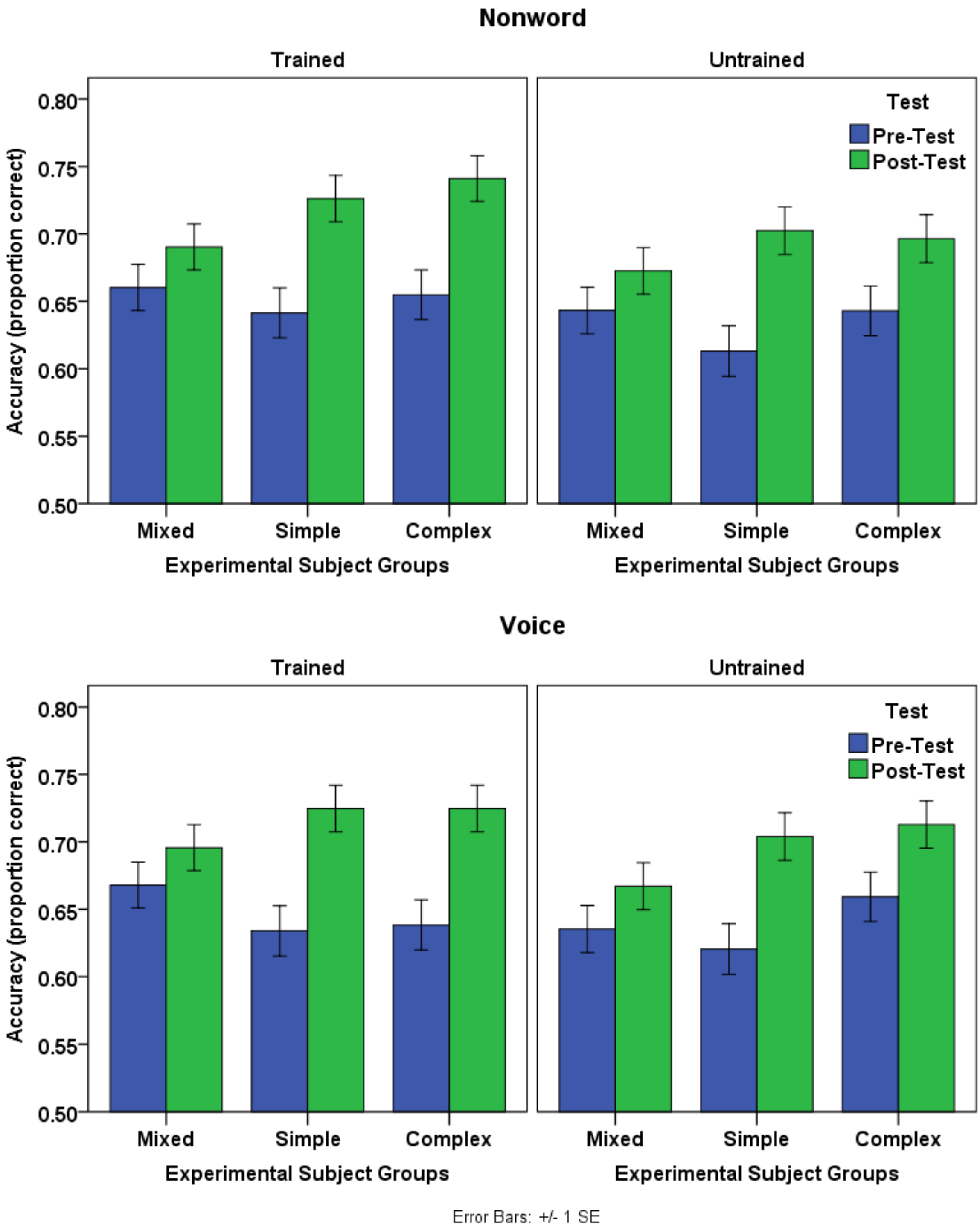


Figure 4. Pre-test and post-test accuracy scores by experimental group and item type.

DISCUSSION AND CONCLUSIONS

The present study has shown that it is possible to improve L2 learners' perception of a difficult vowel contrast through a series of collaborative map tasks performed in pairs. Although we did not manipulate cognitive complexity experimentally, the sequencing of the map tasks in terms of increasing cognitive complexity led to robust gains in discrimination accuracy and speed. Such gains appeared to be consistent across experimental groups receiving exposure to nonwords that differed in linguistic complexity (operationalized as syllable number). In general, it was harder to discriminate the target /iː/-/ɪ/ contrast in tri-syllabic than in monosyllabic nonwords, but experimental groups receiving exposure to only one type of nonword (EG2 and EG3) significantly improved on both types, whereas the group receiving exposure to both types of nonwords appeared to improve the least (albeit not significantly so). We can therefore conclude that linguistic complexity did not significantly affect the effectiveness of the treatment, although the size of the accuracy gains obtained suggest that exposure to monosyllabic nonwords enabled learners to obtain slightly larger gains on all nonword types. The results also showed that learners were able to generalize gains to new items and speakers, indicating that improvement in discrimination accuracy was robust and suggestive of learning having taken place. Overall the findings of the present study suggest that manipulating task design variables to include a focus on phonetic form is an effective pedagogical strategy in generating pronunciation-focused LREs and in improving learners' accuracy in segmental perception while performing an interactive communicative task. It also suggests that exposing learners to simple stimuli in terms of phonological structure in pronunciation tasks may be more beneficial in learning phonological contrasts than exposing them to more complex stimuli. Further research should examine the effectiveness of manipulating task design in pronunciation tasks to include simple-to-complex task sequencing based on linguistic (phonological) complexity in combination with task sequencing based on cognitive complexity.

The current study presents a number of limitations that require further investigation in future research on task-based pronunciation teaching. One crucial issue to be investigated further is the extent to which the treatment effects observed were mainly due to the interactive map tasks or to the familiarization pre-tasks. This would involve recruiting a second control group that would do the pre-tasks, but not the interactive map tasks. It would also be important to include a delayed post-test to test for differential retention of gains in experimental and control groups.

ACKNOWLEDGMENTS

We would like to thank Ingrid Mora-Plaza, Diana Moreira de Oliveira and Natalia Wisniewska for their help in data collection.

ABOUT THE AUTHORS

Joan C. Mora is associate professor in the Department of Modern Languages and Literatures and English Studies in University of Barcelona (UB). His research has examined the role of contextual and individual factors in the development of L2 speech and oral fluency, and the acquisition of L2 phonology.

Contact information: mora@ub.edu, https://joancmora.weebly.com/

Mayya Levkina is assistant professor in the Department of Modern Languages and Literatures and English Studies in University of Barcelona (UB). Her research focuses on the development of different aspects of English as an L2 (e.g. pronunciation, pragmatics) through TBLT as mediated by working memory capacity, attention and other cognitive skills.

Contact information: mayya.levkina@ub.edu

REFERENCES

Best, C., & Tyler, M. (2007) Non-native and second language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 15-34). Amsterdam: John Benjamins.

Cebrian, J. Mora, J. C., & Aliaga-Garcia, C. (2011) Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In Dziubalska-Kołaczyk, K., Wrembel, M., Kul, M. (ed.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010, Volume I* (pp.41-52). Frankfurt am Main: Peter Lang.

Darcy, I. (2017) A psycholinguist walks into a classroom…: A road-map for bridging research and practice in pronunciation teaching. Keynote presentation given at the *9th Annual Conference Pronunciation in Second Language Learning & Teaching* (PSLLT), September 1-2, 2017, University of Utah, Salt Lake City, Utah, USA.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL Teachers' pronunciation "toolbox". In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93-108). Ames, IA: Iowa State University.

Gurzynski-Weiss, L., Long, A. Y., & Solon, M. (Eds.) (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? [Special issue]. *Studies in Second Language Acquisition,* 39(2).

McKinnon, S. (2017). TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declaratives. *Studies in Second Language Acquisition*, 39(2), 287-317.

Mora, J. & Levkina, M. (2017). Task-based pronunciation teaching and research: key issues and future directions. *Studies in Second Language Acquisition, 39*(2), 381-399.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*(1), 81-108.

Parlak, Ö., & Ziegler, N. (2017). The impact of recasts on the development of primary stress in a synchronous computer-mediated environment. *Studies in Second Language Acquisition*, 39(2), 257-285.

Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22,* 27-57.

Robinson, P. (2003). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21, 45-105.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind.* (pp. 243-268). Amsterdam: John Benjamins.

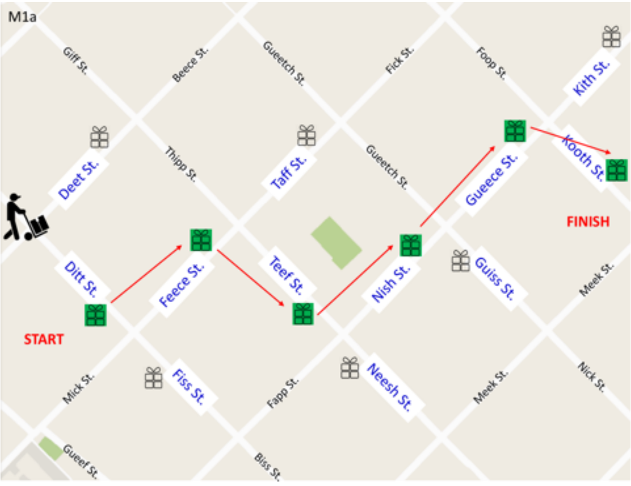
Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *IRAL-International Review of Applied Linguistics in Language Teaching, 45*(3), 161-176.

Saito, K., & Wu, X. (2014). Communicative focus on form and L2 suprasegmental learning: Teaching Cantonese learners to perceive Mandarin tones. *Studies in Second Language Acquisition, 36*, 647-680.

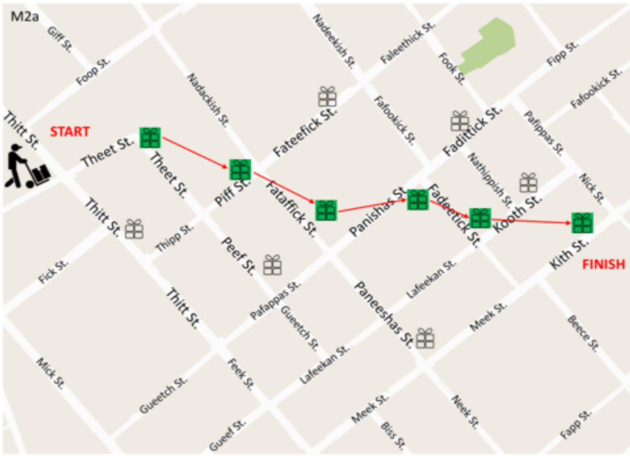
Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition, 39*(2), 347-380.

APPENDIX: SAMPLE MAP TASKS

Simple Map Task



+Complex Map Task



++ Complex Map Task