Studies in Second Language Acquisition (2023), **45**, 571–585
doi:10.1017/S0272263122000213

RESEARCH REPORT

Revisiting the moderating effect of speaker proficiency on the relationships among intelligibility, comprehensibility, and accentedness in L2 Spanish

Amanda Huensch[1]* and Charlie Nagle[2]

[1]University of Pittsburgh, Pittsburgh, PA, USA; [2]Iowa State University, Ames, IA, USA
*Corresponding author. E-mail: amanda.huensch@pitt.edu

(Received 20 October 2021; Revised 14 April 2022; Accepted 25 April 2022)

Abstract

This report examines the potential impacts of task and proficiency on listener judgments of intelligibility, comprehensibility, and accentedness in L2 Spanish. This study extends Huensch and Nagle [*Language Learning, 71,* 626–668, (2021)], who explored the partial independence among the global speech dimensions for speech samples taken from a picture narrative task. Given that the type of speaking task used to elicit speech samples has been shown to impact the strength of the linguistic features contributing to the global speech dimensions and to explore the impact of task on the relationships among the dimensions, the current study followed the same procedure as Huensch and Nagle but employed a task in which participants responded to a prompt based on NCSSFL-ACTFL Can-Do Statements. The speech samples were elicited from instructed L2 Spanish learners of varying proficiency ($n = 42$) and were rated by a group of native-speaking Spanish listeners ($n = 80$) using Amazon Mechanical Turk. In general, the results were consistent with those reported in the initial study indicating a significant, positive, and consistent relationship between comprehensibility and intelligibility and a null relationship between accentedness and intelligibility. The limited differences between the studies' findings are discussed considering the potential impact of task.

Introduction

Evidence for the partial independence of the global speech dimensions of intelligibility, comprehensibility, and accentedness has resulted in a shifting of L2 pronunciation teaching and learning goals away from nativeness principles—achieving nativelike pronunciation using accent reduction—toward intelligibility principles whose focus is on achieving understandable pronunciation (Derwing & Munro, 2015; Levis, 2005, 2020).[1]

[1]Intelligibility and comprehensibility in the current study are defined in line with the conceptualizations of Derwing and Munro: They are related, yet distinct constructs. In other words, the current work is conducted

The limited work exploring the relationships among all three of these global speech dimensions (e.g., Derwing & Munro, 1997; Huensch & Nagle, 2021; Jułkowska & Cebrian, 2015; Munro & Derwing, 1995; Munro et al., 2006; Nagle & Huensch, 2020) has demonstrated stronger and more consistent relationships between intelligibility (the extent to which a listener has understood a speaker's message) and comprehensibility (the ease or difficulty a listener encounters trying to understand a speaker's message) in comparison to intelligibility and accentedness (the strength of a speaker's foreign accent as perceived by a listener). As the ultimate goal of language learning is successful communication of messages, the upshot of these findings is that L2 pronunciation teaching goals ought to focus on improving comprehensibility, as opposed to accentedness, because doing so is more likely to have an impact on intelligibility.

In comparison to the relatively limited number of studies that have incorporated measures of intelligibility, more studies have focused on comprehensibility ratings (e.g., Bergeron & Trofimovich, 2017; Crowther et al., 2015a, 2018; French et al., 2020; Isaacs & Trofimovich, 2012; Isbell et al., 2019; O'Brien, 2014; Saito et al., 2016; Trofimovich et al., 2020). In justifying using comprehensibility ratings as opposed to intelligibility measures, researchers have argued that comprehensibility ratings provide an intuitive way to measure the subjective listener experience of processing difficulty, mirroring real-world applications of such judgments (Crowther et al., 2015a; Trofimovich et al., 2020). Additionally, comprehensibility ratings using Likert or sliding scales are relatively quicker and easier to obtain than intelligibility measurements, which typically involve transcription tasks. Nevertheless, if comprehensibility is to be used as a proxy for intelligibility, then it is important to gain a better understanding of the factors that influence the variability of the strength of the intelligibility-comprehensibility relationship.

Beyond the paucity of work incorporating intelligibility measures, our understanding of the strength of the relationships among these global speech dimensions is additionally limited by the fact that most research in this area has relied on a single type of speaking task (i.e., picture narrative) as well as speech data from relatively advanced speakers of L2 English. Huensch and Nagle (2021) sought to contribute to this line of research by including measures of all three speech dimensions and by investigating the speech of instructed learners of L2 Spanish of varying proficiency; however, they used a picture narrative task to elicit speech data. The current study tested the generalizability of these findings by modifying the speaking task to better understand the influence of task on moderating the strength of the relationships among the global speech dimensions, and whether and how proficiency impacts the strength of those relationships.

Relationships among the global speech dimensions

Previous studies incorporating measurements of intelligibility, comprehensibility, and accentedness have generally reported stronger relationships between intelligibility and comprehensibility than between intelligibility and accentedness, but they have also documented substantial interlistener variability in the strength of the relationships (e.g., Derwing & Munro, 1997; Munro & Derwing, 1995). For instance, Munro and Derwing (1995) reported that for 15 of their 18 listeners there was a

within a paradigm that treats intelligibility and comprehensibility as separate constructs and not simply methodologically differently operationalized.

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

significant correlation between comprehensibility and intelligibility whereas that was true for only five listeners for accentedness and intelligibility (p. 86). Similar findings were reported in Jułkowska and Cebrian (2015) where statistically significant correlations were found between comprehensibility and intelligibility for 15 of 18 listeners (ranging in strength from .667 to .825) whereas for accentedness and intelligibility, the same was true for only five listeners with r values ranging from .099 to .686 (p. 224).

A related line of work has examined linguistic predictors of comprehensibility and accentedness. In general, accumulated findings indicate that both phonological and lexicogrammatical features contribute to comprehensibility and accentedness judgments. However, different features have been shown to map onto each listener-based dimension (Trofimovich & Isaacs, 2012), and even among statistically significant features, some (e.g., word stress) seem to be far better predictors than others (Isaacs & Trofimovich, 2012). Furthermore, when features are bundled into factors, the weights of these factors differ depending on the listener-based construct under consideration. Phonological features tend to be more strongly associated with accentedness than with comprehensibility, whereas for lexicogrammatical features, the opposite is true, insofar as they show a stronger relationship with comprehensibility (Saito et al., 2017). Since these baseline studies, a large body of work has begun to examine the factors that could moderate these relationships. In this study, we focus on two: speaker proficiency and task.

Proficiency as a moderator of the relationship among intelligibility, comprehensibility, and accentedness

Huensch and Nagle (2021), a conceptual replication of Derwing and Munro (1997) and Munro and Derwing (1995), explored the relationships among the three global speech dimensions in L2 Spanish and investigated the potential impact of speaker proficiency on the relationships. Their motivation for focusing on proficiency stemmed from differences in those studies regarding the strengths of the relationships among the speech dimensions that were potentially attributable to differences in proficiency between the speaker samples. Huensch and Nagle (2021) hypothesized that the impact of proficiency might be more evident at the higher and lower ends of the proficiency continuum (in comparison to values in the middle) resulting in a curvilinear relationship. In their study, speech samples were elicited from 42 instructed L2 learners of Spanish of varying proficiency using a picture narrative task. Two utterances per speaker were extracted from the beginning of the narratives and used as stimuli in an online transcription and rating task using Amazon Mechanical Turk (AMT). Eighty native speakers of Spanish completed the AMT task. These listeners were recruited from five countries representing the dialect regions learners reported being most exposed to (Argentina, Colombia, Mexico, Spain, Venezuela). Results from the mixed-effects model analysis indicated a significant positive relationship between intelligibility and comprehensibility (consistent across listeners), such that speech rated as one standard deviation above the mean was twice as likely to be perfectly intelligible. In contrast, accentedness was not a statistically significant predictor of intelligibility. Huensch and Nagle also found a significant positive relationship between comprehensibility and accentedness, but this relationship varied significantly across listeners. When proficiency was incorporated into the models, the findings indicated that proficiency did not impact the strength of the relationship between intelligibility and

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

comprehensibility. In contrast, proficiency did have an impact on the strength of the relationship between comprehensibility and accentedness, such that there was a weaker relationship between these two global speech dimensions in higher proficiency speakers.

While these findings contributed to a better understanding of the relationship among these global speech dimensions and the impact of proficiency on those relationships, speech samples were elicited using the same picture narrative as Munro and Derwing (1995) and Derwing and Munro (1997). This methodological choice was desirable to provide a point of comparison when exploring whether findings generalized to L2 Spanish learners of varying proficiency, but it means that findings are still limited to the same type of picture narrative task that much of the previous literature in this area has relied on (Crowther et al., 2015a).

Task effects on measurements of comprehensibility and accentedness

In a series of studies, Crowther and colleagues (Crowther et al., 2015a, 2015b; Crowther et al., 2018) investigated factors contributing to variation in how rated linguistic features of phonology and fluency (e.g., intonation, speech rate) and lexicon, grammar, and discourse (e.g., lexical appropriateness, grammatical accuracy) contributed to predicting comprehensibility and accentedness ratings. Particularly relevant to the current study, Crowther et al. (2018) explored speaking task effects. In addition to a picture narrative task, they employed two speaking tasks selected to represent real-world assessment contexts of their speaker sample: the IELTS long-turn speaking task and the TOEFL iBT integrated task. Using speech samples from 60 L2 English learners from multiple L1 backgrounds who were rated by 10 experienced L1 English listeners, these studies provided evidence that speaking task indeed impacts how linguistic features map onto comprehensibility and accentedness ratings. In line with previous work, for the picture narrative task, both pronunciation and lexicogrammar features were associated with comprehensibility whereas pronunciation features were associated with accentedness. However, a novel finding was that in the IELTS and TOEFL tasks, accentedness became increasingly associated with lexicogrammar features, leading the authors to observe that "linguistic distinctions between accentedness and comprehensibility were thus clearest in the picture task" (p. 454). Nevertheless, correlation analyses indicated that ratings were strongly related across the three tasks (picture = .80, IELTS = .79, TOEFL = .74, p. 450). Finally, although the effects were small, task appeared to systematically impact the ratings such that in the picture narrative task speakers were rated as less accented but also less comprehensible when compared to ratings for the IELTS task.

Crowther et al. (2018) hypothesized that these findings might be, in part, explained by task familiarity and flexibility both from the speakers' and listeners' perspectives. For instance, regarding the picture narrative task, when listeners are familiarized with the story prior to the experimental task, they are likely to establish expectations for what they hear and how it is presented. At the same time, speakers are constrained by these expectations and therefore have less flexibility in the content they provide such that successfully completing the picture narrative task requires using certain vocabulary and following certain narrative conventions. An extension of this is that utterances extracted from the start of a picture narrative task would also likely be quite similar in their linguistic content and structure such that listeners would encounter many comparable utterances. Trofimovich et al. (2020) offered similar explanations in their

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

study exploring how inter-interlocutor comprehensibility ratings evolve over time during dialogic interactions. They discussed how both task and experience with a speaker's speech could potentially influence comprehensibility. For instance, if listeners are familiar with the content (or listening to content where there is strong expectation about what will be uttered), as they would be in a picture narrative, then their processing resources might be freed up to pay more attention to how the speech potentially deviates from their expectations, thus negatively impacting comprehensibility. Put another way, when the speaker does not produce what the listener expects, then the listener must deploy additional resources to process that mismatch, which could lower comprehensibility.

In sum, while much has been learned about global dimensions of L2 speech, the evidence has primarily come from a single data source (i.e., picture narratives) in a single target language (i.e., English), which limits the generalizability of findings. Additionally, while several studies have contributed important findings to the field by examining tasks effects for comprehensibility and accentedness ratings, those studies have not incorporated intelligibility measures. Therefore, it is unknown how speaking task might impact intelligibility-comprehensibility relationships, among others, which is an important question if comprehensibility is going to continue to be used as a proxy for intelligibility in L2 pronunciation research.

Research Questions and Predictions

The current study is an extension of Huensch and Nagle (2021), who investigated the potential impact of speaker proficiency on relationships among the three global speech dimensions in L2 Spanish using a picture narrative task. Motivated by prior work demonstrating variability in strength among the global speech dimensions of accentedness, comprehensibility, and intelligibility as well as task impacts on listener ratings, the current study followed the same methodological procedures as Huensch and Nagle but modified the speaking task variable. The research questions were as follows:

1. To what extent are intelligibility, comprehensibility, and accentedness related to one another in L2 Spanish speech elicited using a prompted response task?
2. To what extent does proficiency affect relationships among intelligibility, comprehensibility, and accentedness in L2 Spanish?

Huensch and Nagle (2021) found a significant positive relationship between intelligibility and comprehensibility that was consistent across listeners but no statistically significant relationship between accentedness and intelligibility. They also found a significant positive relationship between comprehensibility and accentedness that varied significantly across listeners. The picture narrative task they employed likely had a positive impact on intelligibility because it allowed listeners to have strong preconceived expectations about what they would here. In contrast, these expectations likely had a negative impact on comprehensibility ratings because any mismatches between expectations and actual productions might have required the deployment of additional processing resources. The current study employed a speaking task in which participants responded to a prompt based on NCSSFL-ACTFL Can-Do Statements that, like the IELTS task used in Crowther et al. (2018), allowed speakers more flexibility in choosing what to talk about and how to do so. Therefore, we might predict that in the

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

current study overall intelligibility will be somewhat lower, but comprehensibility would be higher. Because listeners in the current study must concentrate on understanding what the speaker is saying, as opposed to being able to determine what the speaker was saying based on expectation, we might predict an even stronger alignment between intelligibility and comprehensibility. In other words, in open-ended speech, when all the listener has to rely on is what the speaker is saying, then comprehensibility might be a very good representation of intelligibility. Regarding accentedness ratings, although the findings from Crowther et al. (2018) might suggest an impact of task, effect sizes were minimal. In terms of the current study, then, these previous findings might allow us to predict that the relationship between accentedness and the other constructs would remain relatively unaffected. Speaker flexibility in choosing what they say and lowered listener expectation in terms of content suggest we make similar predictions regarding research question 2 that incorporates learner proficiency. Huensch and Nagle found that proficiency did not appear to impact the strength of the relationship between intelligibility and comprehensibility, but it did have an effect on the relationship between comprehensibility and accentedness such that the strength of the relationship weakened as proficiency increased. We predicted similar findings in the current study.

Method

For comparison, Table 1 provides a summary of the similarities and differences between the Method of Huensch and Nagle (2021) and the current study.

Participants

Speakers

Speakers included the same 42 instructed L2 Spanish learners from Huensch and Nagle (2021) who were recruited from first- through fourth-year Spanish courses at two institutions in the United States. Participants were all native speakers of English who

Table 1. Summary of Method differences between Huensch and Nagle (2021) and the current study

	Huensch and Nagle (2021)	Current Study
Speakers	Spanish L2 learners ($n = 42$)	Same
Listeners	Spanish NSs ($n = 80$)	Different NS listeners ($n = 80$)
Speaking task	Hunter story	Prompted response
Rating task	Amazon Mechanical Turk	Same

Table 2. Summary of listener characteristics

	M (SD)	Range
Age	35.33 (9.73)	19–62
Age of onset L2 English	7.11 (3.83)	0–22
Self-reported global English proficiency*	6.86 (1.41)	2.25–9.00
Percent daily English use	15.16 (12.93)	0–60
Familiarity L2 Spanish**	6.39 (2.20)	1–9
L2 teaching experience:	Yes: 17	No: 63

*The proficiency scale ranged from 1–9 (1 = *extremely poor*, 9 = *extremely proficient*).
**The familiarity scale ranged from 1–9 (1 = *not at all familiar*, 9 = *extremely familiar*).

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

reported using English most of the time during the week (M = 94%, SD = 6%). The participants represented a range of proficiencies as indicated by their scores on an Elicited Imitation Test (EIT): M = 55.88 (SD = 26.48), 95% CI [47.63, 64.13], Range 17–106 (out of 120). Four native speakers were also recruited to provide speech samples to ensure that listeners understood the ratings scales and tasks.

Listeners

Listeners included 80 NSs of Spanish recruited from multiple countries (e.g., Mexico, Spain) using AMT. Listeners were recruited using the same IP address filters as those in Huensch and Nagle (2021) in an effort to represent the major dialect zones of instructors at the two institutions of the speakers. The final sample included listeners from Spain (n = 40), Venezuela (n = 20), Mexico (n = 10), Colombia (n = 7), and Argentina (n = 3). The goal was not to construct a set of homogenous listeners, but rather to have speakers evaluated by a range of listeners representing the varieties the speakers had been exposed to and might interact with in the future. Table 2 provides a summary of listener characteristics.

Materials and Procedure

Here, we give a brief overview of our materials and procedure. For complete methodological details, see Huensch and Nagle (2021) whose materials, experimental and coding protocols, data, and analysis code are publicly available at https://osf.io/4j5cr/. Data and analysis code for the current study are available at https://osf.io/4p7r8/.

Speaking task

Speakers completed a speaking task modeled on the NCSSFL-ACTFL Can-Do Statements in which they responded to the following prompt: *Describe un lugar que hayas visitado o que te interese visitar y explica por qué fuiste o por qué quieres ir a ese lugar* "Describe a place you have visited or are interested in visiting and explain why you went there or why you might want to go to this place." Participants were given time to think about their responses and were asked to speak for approximately 1 minute. Two utterances minus any initial hesitation markers were extracted from the start of each response to be used for the rating and transcription task. Utterances from the L2 speakers in the current study were on average 9.3 words (SD = 3.7) with a range of 4–17 words.

AMT rating task

The Human Intelligence Task (HIT) deployed to AMT workers included: (1) a consent form and information about the rating task, (2) a listener background questionnaire, (3) instructions and four practice items, and (4) the experimental rating task. For each item in the rating task, the listeners first heard an utterance one time. Then, the rating interface became active, and listeners had 45 seconds to transcribe the utterance and rate its accentedness and comprehensibility on 100-point sliding scales whose end points were marked with *muy díficl de entender / muy fácil de entender* ("very difficult to understand" / "very easy to understand") for comprehensibility and *acento extranjero muy marcado / ningún acento extranjero* ("very strong foreign accent" / "no foreign accent") for accentedness. Ratings were recorded as numerical values on a 100-point scale (but listeners did not see the numbers).

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

Language background questionnaires

The L2 speakers completed a language background questionnaire to gather basic demographic information about themselves and their language learning experiences. They were also asked about the varieties of Spanish spoken by their instructors, and this information guided the AMT task deployment. The native speaker listeners completed a similar background questionnaire and were asked about their experience and familiarity with L2 Spanish speech.

Scoring and analysis

Intelligibility coding

Each of the utterances extracted from the speakers' open-ended responses was transcribed in CLAN (MacWhinney, 2000) and checked by a second member of the research team. Listener transcriptions from the AMT HIT were compared to the researchers' transcriptions to determine an intelligibility score for each utterance computed as the percentage of words transcribed accurately. Trivial transcription differences (e.g., grammatical regularizations such as transcribing el_{MASC} $día_{\text{MASC}}$ "the day" when the speaker said la_{FEM} $día_{\text{MASC}}$, spelling mistakes such as *aveces* for *a veces* "sometimes") were not considered errors.

Analysis

We adopted the same analytical approach used in Huensch and Nagle (2021). First, we examined the reliability of the comprehensibility and accentedness ratings using two-way, consistency, average-measure intraclass correlation coefficients (ICC). For comprehensibility, ICC = .98 [.98, .99] and for accentedness, ICC = .98 [.97, .99], suggesting that listeners were highly consistent in their use of the two scales. Next, we inspected the distribution of the three scores. The intelligibility data showed extreme left-skew, with most values occurring at 1 (i.e., perfect intelligibility). This amount of skew would have affected the normality of model residuals. We therefore transformed intelligibility scores into a new binary measure, where scores < 1 were coded as 0, or not (perfectly) intelligible, which aligns with the same transformation applied to the intelligibility data in Huensch and Nagle (2021). We then fit a logistic mixed-effects model to the binary intelligibility outcome. Comprehensibility scores were reasonably distributed throughout the 100-point scale, which indicated that there would be no issue with proceeding with the linear mixed-effects models.

We included comprehensibility and accentedness as predictors of intelligibility, alongside the listener-level covariates identified as potentially impactful in Huensch and Nagle (2021): biological age, age of onset of L2 English, self-reported percent daily English use, self-reported global English proficiency, self-reported familiarity with L2 Spanish speech, and a categorical predictor to account for whether listeners had L2 teaching experience. We also included length of utterance, in syllables, as an utterance-level covariate. All continuous predictors were standardized. With respect to the random effects structure of the model, we adopted by-speaker and by-listener random intercepts, testing by-listener random slopes for focal predictors when the corresponding fixed effect reached significance. Testing by-listener random slopes allowed us to estimate between-listener variation in the relationship between the focal predictor and utterance-level intelligibility. We adopted a similar procedure for probing the relationship between comprehensibility and accentedness. We fit a linear mixed-effects model to the comprehensibility data with accentedness as our focal predictor, including the

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

same covariates as above and testing the same random effects. We also included intelligibility as a covariate so that we could estimate the relationship between accentedness and comprehensibility after controlling for the intelligibility of the utterance.

After building these primary models, we tested interactions between comprehensibility and accentedness and proficiency (i.e., participants' EIT score) in the intelligibility model and an interaction between accentedness and proficiency in the comprehensibility model. We examined proficiency as both a linear and quadratic moderating variable, on the view that the moderating effect of proficiency on the relationship between the listener-based constructs might not be linear. For the linear mixed-effects model fit to the comprehensibility outcome variable, we checked the following assumptions: normality of residuals using QQ plots, linearity by plotting fitted values against residuals, and multicollinearity by computing variance inflation factors. Unless otherwise noted, models passed these tests.

Results

As displayed in Figure 1 and mentioned in the preceding text, the intelligibility data were heavily left-skewed, the comprehensibility data showed a relatively even distribution throughout the 100-point scale, and the accentedness data were moderately right-skewed. Descriptive statistics confirmed this trend: for intelligibility, $M = .91$ (.14); for comprehensibility, $M = 58.80$ (29.02); for accentedness, $M = 29.36$ (24.64). Overall, then, it would be fair to characterize the utterances as highly intelligible, moderately comprehensible, and strongly accented.

Interrelationships among the listener-based constructs

The logistic mixed-effects model fit to the binary intelligibility data showed a significant relationship between comprehensibility and intelligibility (*Odds Ratio* = 2.05, 95% CI = [1.86, 2.26), $p < .001$), whereas the relationship between accentedness and intelligibility did not reach significance (*Odds Ratio* = 1.02, 95% CI = [0.93, 1.13], $p = .62$). The odds ratio of 2.05 for comprehensibility indicates that, on average, utterances that were 1 *SD* more comprehensible (where 1 *SD* corresponds to 29.02 units on the 100-point comprehensibility scale) were twice as likely to be intelligible. The marginal R^2 was .20, which indicates that the fixed effects accounted for approximately 20% of variance in intelligibility, and the conditional R^2, which includes the random effects, was .46, indicating that the fixed and random effects together explain 46% of the

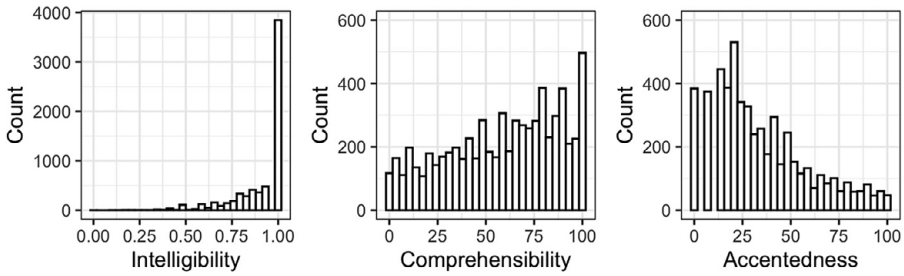

Figure 1. Distribution of intelligibility, comprehensibility, and accentedness scores.

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

variance in intelligibility. By-listener random slopes for comprehensibility did not improve the fit of the model ($\chi^2(2) = 4.22$, $p = .12$), suggesting that the relationship between comprehensibility and intelligibility was consistent across listeners.

The linear mixed-effects model fit to the comprehensibility data revealed a significant positive relationship between accentedness and comprehensibility, after controlling for intelligibility: *estimate* = 10.73, 95% CI = [9.43, 12.03], $p < .001$. This estimate demonstrates that utterances rated as 1 *SD* less accented (where 1 *SD* for accentedness corresponds to 24.64 units on the 100-point scale) tended to be judged as 10.73 units more comprehensible. Integrating by-listener random slopes for accentedness improved the fit of the model: $\chi^2(2) = 313.12$, $p < .001$. Thus, whereas the relationship between comprehensibility and intelligibility was consistent across listeners, the relationship between accentedness and comprehensibility varied considerably. The marginal R^2 of this model was .33 and the conditional R^2 was .64. Thus, the fixed effects accounted for approximately 33% of the variance in comprehensibility versus 64% for the fixed and random effects together.

Overall, then, these initial models showed a strong and stable relationship between comprehensibility and intelligibility and a strong but variable relationship between comprehensibility and accentedness. Considering the effect size benchmarks proposed by Plonsky and Ghanbar (2018), where $R^2 < .20$ is small, $.20 < R^2 < .50$ is medium, and $.50 < R^2$ is large, these models could be considered in the small (intelligibility) to medium (comprehensibility) range.

Effect of proficiency on interrelationships among the listener-based constructs

We included proficiency as a moderating variable by generating interactions with the focal predictors in each model: for intelligibility, proficiency × comprehensibility and proficiency × accentedness, and for comprehensibility, proficiency × accentedness. We used the poly function to generate orthogonal linear and quadratic terms, and we used likelihood ratio tests to determine if the more complex model with the proficiency interactions significantly improved model fit over a simpler, predecessor model (described in the previous section) that did not include those interactions.

The interaction model for intelligibility was a marginally better fit than the baseline model: $\chi^2(5) = 12.82$, $p = .03$. Interestingly, when we tested a simpler interaction model, including only linear proficiency in interaction with the focal predictors, that model did not prove to be an improvement over baseline: $\chi^2(2) = 2.59$, $p = .27$. This finding indicates that the quadratic moderator was the primary driver of the modest improvement in model fit. Neither of the linear interactions were statistically significant, but both of their quadratic counterparts were. The odds ratio for the quadratic proficiency × comprehensibility interaction was greater than 1 (*Odds Ratio* = 1.11, 95% CI = [1.02, 1.20], $p = .01$), which shows that the relationship between comprehensibility and intelligibility was slightly stronger at the proficiency extremes (i.e., in speakers of lower and higher proficiency). Conversely, the odds ratio for the quadratic proficiency × accentedness interaction was less than 1 (*Odds Ratio* = 0.91, 95% CI = [0.84, 0.97], $p = .01$), which indicates that the relationship between accentedness and intelligibility was weaker at both lower and higher proficiency levels. It bears repeating, however, that the overall relationship between accentedness and intelligibility was not significant. Thus, the significant quadratic interaction has two interpretations: (1) at certain proficiency levels, the relationship between accentedness and intelligibility could be significant, but those

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

levels would likely be extreme and not attested in most speakers and (2) there could be significant differences in the relationship between accentedness and intelligibility in speakers of varying proficiency (i.e., the accentedness-intelligibility slope estimate at proficiency = –1 *SD* could be different from the slope estimate at proficiency = +1 *SD*) despite a nonsignificant overall finding (i.e., each slope may not be significantly different from zero). Furthermore, the marginal R^2 of the interaction model was .21, which represents a negligible 1% improvement over the baseline model ($R^2 = .20$). Thus, it would be fair to say that the statistically significant improvement in model fit was not practically significant. Put another way, relationships between comprehensibility and intelligibility and accentedness and intelligibility do not appear to vary much at all as a function of speaker proficiency.

For comprehensibility, a model with a linear proficiency × accentedness interaction was an improvement over the baseline model ($\chi^2(1) = 34.14$, $p < .001$), but a model with a quadratic interaction did not result in any additional improvement ($\chi^2(2) = 5.66$, $p = .06$). The significant negative coefficient for the interaction term (*estimate* = –1.44, 95% CI = [–1.93, –0.96], $p < .001$) shows that the relationship between accentedness and comprehensibility became slightly weaker in speakers of higher proficiency. Put another way, accentedness seems to be more strongly aligned with comprehensibility at lower proficiency levels. Again, however, considering the 100-point comprehensibility scale and the magnitude of the baseline accentedness estimate, which was 11.39 in the updated model (95% CI = [10.09, 12.69], $p < .001$), the effect of proficiency on the relationship between accentedness and comprehensibility was relatively small. This fact is also confirmed by the marginal R^2 of the interaction model, which remained .33, the same as the baseline model. Thus, as was the case for the intelligibility model, the relationship between accentedness and comprehensibility does not appear to vary with speaker proficiency, at least not in a practically significant way.

Discussion

In this study, we found a significant, positive, and consistent relationship between comprehensibility and intelligibility and a null relationship between accentedness and intelligibility. We also found a significant positive relationship between accentedness and comprehensibility, but that relationship varied significantly across listeners. As shown in the top portion of Table 3, these findings closely align with those reported in Huensch and Nagle (2021). In fact, most coefficients were a near exact match across the studies, which suggests that task had very little effect on the relationships between the listener-based constructs. The only coefficient that changed slightly was the estimate of the relationship between accentedness and comprehensibility, which was slightly smaller in the present study than in Huensch and Nagle (2021). This shrinkage, albeit modest (see the substantial overlap in the 95% CIs), suggests that there is a somewhat weaker relationship between accentedness and comprehensibility when speakers have complete freedom in choosing what grammar and vocabulary to use and when listeners have less concrete expectations about the content of the speech sample. Perhaps then, when listeners have strong expectations about what a speaker will say and the language they will use to say it, they can allocate attention toward the way in which the speaker communicates the information rather than focusing on what they are trying to communicate. As a result, if a speaker does not produce what the listener expects additional processing resources might be required to address the mismatch, which

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

Table 3. Comparison of results: Huensch & Nagle (2021) vs. current study

	Huensch & Nagle (2021) Task: Picture narration	Current Study Task: Prompted response
Descriptive statistics		
Intelligibility	.93 (.12)	.91 (.14)
Comprehensibility	55.62 (29.01)	58.80 (29.02)
Accentedness	30.36 (24.65)	29.36 (24.64)
Baseline models		
Comprehensibility-intelligibility	*Odds Ratio* = 2.07* 95% CI = [1.87, 2.29]	*Odds Ratio* = 2.05* 95% CI = [1.86, 2.26]
Accentedness-intelligibility	*Estimate* = 1.01 95% CI = [0.91, 1.11]	*Estimate* = 1.02 95% CI = [0.93, 1.13]
Accentedness-comprehensibility	*Estimate* = 11.53* 95% CI = [10.23, 12.83] Random slopes *SD* = 5.10*	*Estimate* = 10.73* 95% CI = [9.43, 12.03] Random slopes *SD* = 5.05*
Proficiency models		
Comprehensibility-intelligibility	*na*	Quadratic moderator *Odds Ratio* = 1.11 95% CI = [1.02, 1.20] ΔR^2 = 0.01 (1% variance)
Accentedness-intelligibility	Quadratic moderator *Odds Ratio* = 0.91 95% CI = [0.84, 0.99] ΔR^2 = .05 (5% variance)	Quadratic moderator *Odds Ratio* = 0.91 95% CI = [0.84, 0.97] ΔR^2 = 0.01 (1% variance)
Accentedness-comprehensibility	Linear moderator *Estimate* = −0.83 95% CI = [−1.38, −0.28] ΔR^2 = .00 (0% variance)	Linear moderator *Estimate* = −1.44 95% CI = [−1.93, −0.96] ΔR^2 = .00 (0% variance)

could explain a somewhat stronger comprehensibility-accentedness link for the picture narration samples than for the prompted response samples.

In terms of the moderating effect of proficiency on the relationships between the listener-based constructs, again, findings for the prompted response task closely align with findings for the picture narration task (see the lower portion of Table 3). Huensch and Nagle did not find that proficiency affected the relationship between comprehensibility and intelligibility, whereas in this study we did. However, it is important to bear in mind that this effect was very small, explaining less than 1% of the variance in the intelligibility data. Therefore, despite differences in what reached statistical significance across the two studies, the practical significance of the findings is clear: In both studies, proficiency had very little impact on the relationship between comprehensibility and intelligibility. The same could be said of the effect of proficiency on the relationship between accentedness and intelligibility. Despite reaching statistical significance in both reports, the amount of variance that the interaction term explained was negligible in the current study (1%) and very modest in the previous study (5%). Thus, it would be fair to say that proficiency seems to have very little impact on the accentedness-intelligibility relationship irrespective of task type. Lastly, although integrating proficiency interactions into the comprehensibility model improved model fit, the additional variance in comprehensibility that those terms explained was very small (< 1%). The tentative conclusion that can be reached, then, is that proficiency has little to no impact on the relationships between the listener-based dimensions, which also appear to be consistent across speaking tasks.

https://doi.org/10.1017/S0272263122000213 Published online by Cambridge University Press

Conclusion

The current study's extension of Huensch and Nagle (2021) provides additional evidence for the partial independence of the global speech dimensions of intelligibility, comprehensibility, and accentedness. Importantly, it lends further support to the pedagogical focus on comprehensibility given its stronger and more consistent relationship to intelligibility in comparison to accentedness. Furthermore, the findings indicated a limited effect of speaking task in moderating the strength of the intelligibility and comprehensibility relationship. One consideration for future work relates to the fact that the current study included an intelligibility measure whereas many previous studies only included ratings of accentedness and comprehensibility. This raises an interesting question about whether having raters transcribe the speech might influence ratings of comprehensibility and/or accentedness and thus potentially the strength of their relationship as well. For instance, if a listener is unable to complete the transcription task, this likely indicates to them potential difficulties related to comprehensibility whereas in cases where the transcription task is easily accomplished this might suggest ease in processing. As suggested by Huensch and Nagle (2021), the inclusion of an intelligibility transcription task might explain the relatively consistent relationship between intelligibility and comprehensibility across listeners. This begs the question of whether the strength of the accentedness/comprehensibility relationship might vary depending upon whether or not a transcription task is included (i.e., depending on the methodological characteristics of the research design). For instance, Derwing and Munro (1997), which included intelligibility, reported a mean correlation of $r = 0.45$ (p. 7) whereas Saito et al. (2016) and Isbell et al. (2019), who did not include intelligibility, reported correlation coefficients of $r = 0.89$ (p. 226) and $r = 0.92$ (p. 36), respectively. Future work should explore the potential impact of these methodological differences. Additionally, as noted by Saito (2021), a fruitful avenue for future meta-analytic work examining the global speech dimensions would be the inclusion of task as part of a moderator analysis, given the growing number of primary studies.

Another interesting avenue for future work will be considering how the complexity and predictability of the speaking sample affect listener-based ratings and the linguistic variables that predict them. For instance, when listeners can easily predict the content of the sample, either because the speaking task is relatively circumscribed or because they received instructions and images representing what the speakers had to do, they may be able to ascertain what speakers have said even if the speech is difficult to process, in which case intelligibility and comprehensibility (and the features that map onto them) might show a weaker relationship. However, when the message is less predictable, then listeners may need to focus entirely on apprehending what the speaker said, bringing intelligibility and comprehensibility closer in line with one another. Regardless of whether such hypotheses are borne out, much more work is needed on how task characteristics and listener background knowledge interact with the linguistic and stylistic variables that the speaker brings to the table. To that end, experimental studies that manipulate those variables could prove especially illuminating.

Acknowledgments. This work was funded by a University of South Florida Creative Scholarship Grant and a University of South Florida Nexus Initiative Award to the first author and an Iowa State University Social Sciences Seed Grant to the second author. We would like to thank the participants and our research assistants, especially Aneesa Ali and Bianca Pinkerton.

Data Availability Statement. The experiment in this article earned Open Materials and Open Data badges for transparent practices. The materials and data are available at https://osf.io/4j5cr/ and https://osf.io/4p7r8/.

References

Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of accentedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals*, *50*, 547–566. https://doi.org/10.1111/flan.12285

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015a). Does a speaking task affect second language comprehensibility? *Modern Language Journal*, *99*, 80–95. https://doi.org/10.1111/modl.12185

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015b). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, *49*, 814–837. https://doi.org/10.1002/tesq.203

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*, 443–457. https://doi.org/10.1017/S027226311700016X

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*, 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.

French, L. M., Gagné, N., & Collins, L. (2020). Long-term effects of intensive instruction on fluency, comprehensibility and accentedness. *Journal of Second Language Pronunciation*, *6*, 380–401. https://doi.org/10.1075/jslp.20026.fre

Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning*, *71*, 626–668. https://doi.org/10.1111/lang.12451

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*, 475–505. https://doi.org/10.1017/S0272263112000150

Isbell, D. R., Park, O. S., & Lee, K. (2019). Learning Korean pronunciation: Effects of instruction, proficiency, and L1. *Journal of Second Language Pronunciation*, *5*, 13–48. https://doi.org/10.1075/jslp.17010.isb

Jułkowska, I. A., & Cebrian, J. (2015). Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech. *Journal of Second Language Pronunciation*, *1*, 211–237. https://doi.org/10.1075/jslp.1.2.04jul

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 369–377. https://www.jstor.org/stable/3588485

Levis, J. M. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation*, *6*, 310–328. https://doi.org/10.1075/jslp.20050.lev

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111–131. https://doi.org/10.1017/S0272263106060049

Nagle, C., & Huensch, A. (2020). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*, *6*, 329–351. https://doi.org/10.1075/jslp.20009.nag

O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech: L2 learner assessments. *Language Learning*, *64*, 715–748. https://doi.org/10.1111/lang.12082

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *Modern Language Journal*, *102*, 713–731. https://doi.org/10.1111/modl.12509

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*, 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*, 217–240. https://doi.org/10.1017/S0142716414000502

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462. https://doi.org/10.1093/applin/amv047

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*, 905–916. https://doi.org/10.1017/S1366728912000168

Trofimovich, P., Nagle, C. L., O'Brien, M. G., Kennedy, S., Taylor Reid, K., & Strachan, L. (2020). Second language comprehensibility as a dynamic construct. *Journal of Second Language Pronunciation*, *6*, 430–457. https://doi.org/10.1075/jslp.20003.tro

Cite this article: Huensch, A. and Nagle, C. (2023). Revisiting the moderating effect of speaker proficiency on the relationships among intelligibility, comprehensibility, and accentedness in L2 spanish. *Studies in Second Language Acquisition,* **45,** 571–585. https://doi.org/10.1017/S0272263122000213