

Listener perception of appropriateness of L1 and L2 refusals in English

Maria Kostromitina ✉

Northern Arizona University, Flagstaff, USA

<https://orcid.org/0000-0002-2505-7762>

masha@nau.edu

Yongzhi Miao

Northern Arizona University, Flagstaff, USA

<https://orcid.org/0000-0002-1467-5688>

vmiao@nau.edu

Abstract

English has become an international language (EIL) as speakers around the world use it as a universal means of communication. Accordingly, scholars have investigated different aspects of EIL affecting communicative success. Speech scholars have been interested in speech constructs like accentedness, comprehensibility, and acceptability (e.g., Kang et al., 2023). On the other hand, pragmatic researchers have examined lexico-grammatical features of EIL that contribute to first language (L1) English listeners' perceptions of appropriateness in speech acts (e.g., Taguchi, 2006). However, little is known about: a) how appropriateness is perceived by users of EIL of diverse L1s and b) how those appropriateness perceptions are related to lexico-grammatical and phonological features. Therefore, the present study had 184 listeners (L1 = English, Spanish, Chinese, and Indian languages) evaluate 40 speech acts performed by 20 speakers (L1 English and Chinese, 50% each) in terms of appropriateness on a 9-point numerical scale. Results from linear mixed-effects regressions suggested that: a) listener L1 did not contribute to listener ratings and b) speakers' rhythm and lexico-grammatical features (i.e., use of different pragmatic strategies) significantly contributed to listener appropriateness ratings. The findings provide empirical evidence to support the phonology-pragmatics

link in appropriateness perceptions and offer implications regarding the operationalization of English interactional appropriateness.

Keywords: speech perception; appropriateness; pragmatics; phonology

1. Introduction

English nowadays is used as a lingua franca by both first language (L1) and second/foreign language (L2) English users (Eberhard et al., 2022). Given this status of English in international communication, scholars in the fields of L2 speech and pragmatics have been interested in listeners' perception of English use by speakers of diverse English varieties and how those perceptions affect the success of interaction. From the point of view of speech perception, scholars have examined the role of listener perceptions of *accentedness* (i.e., phonological differences from L1 varieties), *comprehensibility* (i.e., ease of understanding), and *acceptability* (i.e., suitability of an utterance in a given communicative context) in intercultural communication and the acoustic elements of speech that contribute to those perceptions. Parallel to these speech constructs, in the field of pragmatics, researchers have approached interlocutors' perceptions of successful interaction from the viewpoint of pragmatic appropriateness (e.g., Al Masaeed et al., 2020; Cunningham, 2017), looking particularly at the use of pragmatic strategies and lexico-grammatical modification devices. This study bridges the gap between the two domains by expanding the definition of *appropriateness* as listeners' perceptions of learners' ability to produce speech acts (i.e., making refusals) in an interpretable and effective way that incorporates not only pragmatic strategies but also acoustic elements of speech, following research in speech perception (see Cunningham, 2017; Kostromitina, in press; van Compernelle, 2014). Although the connection between speech constructs and pragmatics has consistently been made in previous research (e.g., Brazil, 1997; Pickering, 2018), the field still does not have a clear understanding of the exact role of segmental and suprasegmental features of speech in pragmatic perceptions. By taking this integrated approach, the study seeks to clarify the construct of appropriateness in intercultural communication and for further use in English teaching and testing. In addition, while previous studies have explored appropriateness perceptions of L1 English users (e.g., Taguchi, 2006), this study expands the participant population to L2 English users to allow for a more comprehensive understanding of listener perception of appropriateness in the currently globalized world where English is used as a lingua franca.

2. Literature review

2.1. Appropriateness as a pragmatic construct

Pragmatic appropriateness has been one of the central constructs in L2 pragmatic research used to measure interlocutors' pragmatic competence, especially within speech acts. Although common in research, as noted by Dewaele (2008), appropriateness is a "slippery" construct that is hard to define (p. 246). In fact, the operationalization of appropriateness has differed from study to study. For example, Taguchi (2011) defined the construct of pragmatic appropriateness as the degree of politeness, directness, and formality that is suitable for a certain communicative situation. Purpura (2004), in turn, pointed out the latent nature of appropriateness as incorporating nativeness, politeness, as well as cultural norms, preferences, and expectations. In other studies (e.g., Cunningham, 2017; van Compernelle, 2014), appropriateness was defined broadly as listeners' perceptions of learners' ability to produce speech acts in an interpretable and effective way.

Notwithstanding the differences in operationalizing this construct, a small group of studies have attempted to measure L1 English speakers' perceived pragmatic appropriateness and identify criteria that L1 English users consider when rating the appropriateness of pragmatic production. Thus, Alemi et al. (2014) collected 60 Iranian English teachers' ratings of politeness, tactfulness of production, sincerity, and socio-pragmatic and linguistic appropriateness. Similarly, Tajeddin and Alemi (2014) had 50 L2 English teachers rate L2 speakers' responses to six written discourse completion tasks and found that oftentimes listeners considered perceived politeness to be one of the aspects of pragmatic appropriateness. In addition, later studies by Alemi and collaborators (e.g., Alemi & Khanlarzadeh, 2016; Alemi & Motamedi, 2019) added situationally dependent formality and authenticity as criteria used by listeners. L1 English speakers' perceived appropriateness has also been measured in studies via Likert scales (e.g., Cunningham, 2017; Taguchi, 2013).

Despite the growing body of research in pragmatic appropriateness, the majority of work has centered around L1 English perceptions, thus bypassing L2 speakers' perceptions and overlooking potential differences one's cultural and linguistic backgrounds may introduce into their definition of this construct. In addition, little is known about the differences in appropriateness perceptions between L2 speakers of English from different backgrounds.

2.2. Speech perception constructs

In speech perception studies, a dominant line of research has explored listener perceptions of accented speech in terms of constructs like accentedness, comprehensibility

and acceptability, and linguistic features that contribute to these constructs (e.g., Isaacs & Trofimovich, 2012). Research has suggested that diverse linguistic factors, including pronunciation, lexicogrammar, and discourse organization are related to listeners' perceptual judgments of L2 speech (e.g., Gass & Varonis, 1984; Isaacs & Trofimovich, 2012; Kang et al., 2023; Saito et al., 2017). Thus, moving beyond linguistic features, scholars within the domain of L2 speech perception have argued that communicative competence entails more than one's use of lexical and grammatical structures, emphasizing the role of acoustic features of speech in listeners' perceptions of one's production (Bachman & Palmer, 2010; Hymes, 1971). Specifically, Brazil (1997) and later Pickering (2018) suggested that one's use of suprasegmental features (e.g., intonation, prominence, pitch height) may affect the process of meaning making in interaction and thus affect its outcome. This link between suprasegmental or prosodic features of speech has been made in several recent studies that investigated L2 English users' production of speech acts (e.g., Kang et al., 2023; Taguchi et al., 2022). From a perceptual perspective, Sydorenko et al. (2014) examined three L1 English speakers' appropriateness perceptions in L2 requests with a mixed-methods approach, where the raters first listened to the requests produced by L2 speakers, rated them in terms of pragmatic appropriateness, and provided explanations to justify their ratings. This study found that specific intonation patterns were associated with higher or lower appropriateness ratings. Outside of ELF contexts, Herrero and Devis (2020) investigated 30 L1 Spanish speakers' perception of 100 L2 Spanish requests produced by 20 L1 Mandarin speakers. After listener evaluation of the requests regarding politeness, the scholars performed prosodic analysis. Results suggested that certain intonation contours paired with a tonal emphasis on stressed syllables affected listeners' politeness perceptions.

Although research has consistently made a link between phonological features of speech and various listeners' perception constructs like comprehensibility, accentedness, and acceptability, very few studies have made such an association with the pragmatic construct of appropriateness. This is surprising given the theoretical support and emerging empirical evidence regarding the role of phonological features in pragmatic meaning making and meaning interpretation. Additionally, existing scarce appropriateness perception studies have exclusively examined L1 speakers' ratings bypassing L2 speakers' perceptions, similarly to research in L2 pragmatics discussed in the previous section.

2.3. Pragmatic appropriateness and suprasegmental speech features

The sections above introduced the construct of pragmatic appropriateness and emerging interdisciplinary perception research that surrounds this construct in

the domains of L2 pragmatics and speech. An assumption underlying pragmatic appropriateness based on this research is that beyond what to say, how we say it is also important (see Bachman & Palmer, 2010; Hymes, 1971). There are two ways to unpack the “how we say it” part of the statement: a) based on lexico-grammatical features (used interchangeably with pragmatic strategies) and b) based on phonological features. First, the choice of lexical or grammatical devices may change the pragmatic meaning of a given sentence. Studies have revealed different use of pragmatic strategies between English speakers of different L1s (e.g., Krulatz & Dixon, 2020), and between L1 English speakers and L2 speakers (e.g., Taguchi, 2013), regarding the use of direct versus indirect strategies. Direct strategies (e.g., *I don't want to do this*) are often perceived as less ideal than indirect strategies (e.g., *I am sorry, but I have an important meeting on Friday*) in terms of refusals. This means that lexico-grammatical features of the speech acts are related to one's pragmatic competence.

Second, differences in the phonological output (e.g., prosody) may change the intended meaning perceived by the listener. Scholars within the field of both speech perception and pragmatics are no strangers to this argument. For instance, Hirschberg (2017) provided a detailed account of the relationship between prosody and semantic interpretation (i.e., differences in terms of meaning) and discourse phenomena (e.g., old vs. new information). In a similar vein, scholars have illustrated the connection between prosody and pragmatics in Catalan (Prieto & Rigau, 2012) and English (Brazil, 1997; Wharton, 2012) as exemplar language varieties. Common in this line of research is that many scholars have taken a theoretical perspective, and their theory-informed arguments would benefit from empirical support. Moreover, many of the interpretations were made on the beliefs of the L1-speaking communities. The inclusion of speakers of different spoken varieties may reveal a more complex picture of the perception of different English varieties. In the present study, we include both lexico-grammatical features and prosodic features within the same model to predict listener appropriateness ratings of L1 and L2 English.

3. The present study

Overall, two pertinent gaps have been identified at the interface of speech and pragmatics. First, many studies investigated listener perception of appropriateness based solely on the L1 English population. The inclusion of listeners of diverse spoken varieties would likely provide additional insights into how L2 speakers are perceived in a globalized world. Second, many studies made theoretical arguments about speech-pragmatics connection. These arguments would benefit from empirical support that

includes measurable parameters of speech and pragmatics. To address these gaps, the present study was guided by the following research questions:

1. To what extent do listeners of different L1s perceive appropriateness of refusals differently?
2. What is the relationship between perceived appropriateness of refusals and the prosodic and pragmatic properties of refusals?

4. Method

4.1. Participants

4.1.1. Speakers

The present study is based on a corpus collected for a dissertation within the research team. The corpus included both L1 ($N = 34$) and L2 speakers of English ($N = 49$), totaling 83 speakers. The L1 speakers in the corpus were undergraduate students across different majors at the time of the study, with a balanced gender ratio (male:female = 16:18). The L2 speakers were also undergraduate students enrolled in the upper-level Intensive English Program at the university. The L2 sample is skewed in terms of gender ratio (male:female = 6:63) and L1s (Mandarin:Russian = 46:3).

All speakers were asked to produce a total of ten speech acts (refusals:requests = 5:5) via an online questionnaire. Within each speech act, the speakers had three turns in total: a) conversation opening, b) actual speech acts, and c) conversation closings. Only the actual speech acts were included in the present corpus. Overall, the corpus encompassed 830 speech files, totaling about 318 minutes of speech. The present study only used the refusals in the corpus at its initial stage ($N = 415$ files), specifically, two scenarios for the refusals: a) a student being asked to give a class presentation by the professor ($N = 208$ files) and b) a student being asked to substitute for a coworker by their employer ($N = 207$ files). These scenarios were adapted from previously validated discourse completion tasks (Félix-Brasdefer, 2003; Taguchi, 2012) and were deemed appropriate for English users in academic settings.

10 L1 English and 10 L1 Chinese speakers were selected from a collection of speech files from an ongoing project. They were selected based on: a) high intelligibility, b) good recording quality, c) correct interpretation of the prompt, and d) balanced gender ratio. First, intelligibility was assessed by two independent researchers. They listened to each sound file and transcribed them. Only

recordings with an accuracy rate of over 80% for both transcribers were potentially included. Second, during transcription, they flagged recordings with noticeable background noise. These recordings were discarded. Third, recordings that did not follow the instructions were discarded. These included cases where speakers did not make a refusal as instructed. Last, of the remaining, qualified files, random selection was made (controlling for gender ratio), resulting in the final speech files used in the present study ($N = 40$ files).

4.1.2. Listeners

A total of 184 listeners from four L1 backgrounds will be recruited: a) L1 English ($N = 73$), Chinese ($N = 77$), South American Spanish ($N = 27$), and Indian languages ($N = 8$). In the case when they spoke L2 English, their proficiency was high-intermediate. They were all undergraduate students of different majors (19–23 years old). Their genders across different subsamples were similarly skewed, with over 70% identifying as female.

4.2. Materials and instruments

4.2.1. Online questionnaire

An online questionnaire was developed via Qualtrics for data collection. The questionnaire included: a) information sheet and consent form, b) the 40 files embedded and presented in random order, and c) questions exploring listener background.

For each speech file, listeners were asked seven semantic-differential questions, including one question probing appropriateness (i.e., The speaker was . . . ; 1 = *completely inappropriate*, 7 = *completely appropriate*), three questions probing comprehensibility (e.g., The speaker was . . . ; 1 = *difficult to understand*, 7 = *easy to understand*), and three questions probing accentedness (e.g., The speaker . . . ; 1 = *speaks with a foreign accent*; 7 = *speaks with an American accent*). This study only includes the analysis of the appropriateness scale. Besides the seven Likert-scale questions for quantitative data, each speech file was accompanied by an optional comment box where listeners could leave qualitative comments (in their L1s, if they preferred). Many listeners commented with justifications for the scores they assigned. The background questionnaire asked about listeners' demographic information (e.g., gender), education background (e.g., highest degrees, majors), and language background (e.g., proficiency in English and additional languages, if any).

4.3. Data coding and analysis

4.3.1. Coding

All of speech files were coded based on several phonological features and pragmatic strategies. Regarding phonological coding, three features were included: a) tone choice, b) pitch range, and c) space. The rationale for selecting these suprasegmental features was that they were significant predictors of listeners' perceptual judgments of L2 speech (see e.g., Kang, 2010). All coding was completed by two graduate students of applied linguistics, and high intercoder reliability was achieved (see Kostromitina, in press). Agreement reached 92% on the identified tone choices. Discrepancies were resolved via discussion. Below is a detailed description of the coding schemes for phonological features and lexico-grammatical features (i.e., pragmatic strategies).

Tone choice is a measure of intonation and prosody. There are different frameworks of coding tone choice, and the research team followed the model proposed by Brazil (1997). To code for tone choice, the research team first segmented the speech into different tone units, that is, the minimal meaningful units that entail a pitch change. This was done by examining the presence of focus word (that receives prominence). Then, the research team examined the pitch change of the prominent syllable of the focus word and coded for tone choice as follows: a) falling, b) rising, and c) level. Specifically, the research team only coded for the tone choice of the speech act. In the case where there were multiple sentences within the speech act, we only coded for the core sentence that expressed refusal because it conveys the core pragmatic meaning (see discussion for its potential limitation). The coding was performed manually in PRAAT (Boersma & Weenink, 2018).

Pitch range measures the distance between the upper limit and the lower limit of one's pitch contour. There are two different ways of measuring it: a) by examining all of the prominent syllables regarding the maximum and the minimum pitch value and b) by examining all words regarding the maximum and the minimum pitch value. The research team opted for the second option (see discussion for its potential limitation).

Space is a measure of speech rhythm (i.e., the use of strong and weak syllables in combination). It refers to the proportion of prominent words (see Vanderplank, 1993). This was computed after all the prominent words had been identified following this formula (number of stressed words/total number of words).

In terms of pragmatic strategies, the coding scheme was adopted from Krulatz and Dixon (2020) who synthesized previous coding schemes by Beebe et al. (1990) and Salazar Campillo et al. (2009). A total of 11 direct (2) and indirect

(9) strategies were included. The direct strategies comprised: a) bluntness/the use of performative verbs (e.g., *I have to decline*) and b) the use of non-performative statements/negotiation of proposition (e.g., *I can't make it*). The indirect strategies comprised: a) use of plain indirect statements (e.g., *I'm not sure if I'll be able to talk about my paper*), b) expressing wish (e.g., *I wish I could go*), c) provision of reason (e.g., *I've got some assignments to do*), d) statement of regret/apology (e.g., *I am sorry, but ...*), e) provision of alternatives (e.g., *I would love to come at another time*), f) expression of disagreement (e.g., *I'm not sure if this is the best way to help students*), g) statement of principle/philosophy (e.g., *I do not feel confident to teach others*), h) provision of advice (e.g., *perhaps you should ask someone else*), i) avoidance, including hedging (e.g., *unfortunately, ...*), change of topic (e.g., *by the way, what are you doing next week?*), and sarcasm (e.g., *don't have too much fun without me*). Table 1 provides a summary of these strategies. When coding, in the case where there were multiple strategies within the speech act, we only included the first one (see discussion for its potential limitation).

Table 1 Coding schemes for pragmatic strategies adopted from Krulatz and Dixon (2020)

	Strategies	Examples
Direct Strategies	bluntness/the use of performative verbs	<i>I have to decline.</i>
	use of non-performative statements/negotiation of proposition	<i>I can't make it.</i>
Indirect Strategies	use of plain indirect statements	<i>I'm not sure if I'll be able to talk about my paper.</i>
	expressing wish	<i>I wish I could go.</i>
	provision of reason	<i>I've got some assignments to do.</i>
	statement of regret/apology	<i>I am sorry, but ...</i>
	provision of alternatives	<i>I would love to come at another time.</i>
	expression of disagreement	<i>I'm not sure if this is the best way to help students.</i>
	statement of principle/philosophy	<i>I do not feel confident to teach others.</i>
	provision of advice	<i>Perhaps you should ask someone else.</i>
	avoidance: hedging, change of topic, sarcasm	<i>Unfortunately, ...</i> <i>By the way, what are you doing next week?</i> <i>Don't have too much fun without me!</i>

4.3.2. Analysis

The first research question investigated to what extent listeners' appropriateness ratings differ depending on their L1 background. To answer this question, a linear mixed-effects model was computed in R studio, and preliminary analyses were done to ensure that the assumptions pertaining to the model were

met. Fixed effects within the models included listeners' L1. Random effects included listener idiosyncrasy, listener gender, speaker idiosyncrasy, pragmatic scenario (two situations for refusals), speaker L1, and speaker gender:

$$\begin{aligned} \text{Appropriateness} \sim & \text{Listener_L1} + \\ & (1 | \text{Listener_ID}) + (1 | \text{Listener_Gender}) + \\ & (1 | \text{Speaker_ID}) + (1 | \text{Scenario}) + (1 | \text{Speaker_L1}) + (1 | \text{Speaker_Gender}) \end{aligned}$$

The second research question investigated the relationship between perceived appropriateness of refusals in English and the prosodic and pragmatic properties of the refusals. To answer this question, a linear mixed-effects model was computed in R studio, and preliminary analyses were done to ensure that the assumptions pertaining to the model were met. Fixed effects within the models included tone unit (with three levels: falling, rising, and level) and pragmatic features (with nine levels). Random effects included listener idiosyncrasies, listener gender, listener L1, speaker idiosyncrasies, pragmatic scenario, speaker L1, and speaker gender:

$$\begin{aligned} \text{Appropriateness} \sim & \text{Tone_Choice} + \text{Space} + \text{Pitch} + \text{Strategies} + \\ & (1 | \text{Listener_ID}) + (1 | \text{Listener_Gender}) + (1 | \text{Listener_L1}) + \\ & (1 | \text{Speaker_ID}) + (1 | \text{Scenario}) + (1 | \text{Speaker_L1}) + (1 | \text{Speaker_Gender}) \end{aligned}$$

5. Results

5.1. Appropriateness and listener L1

This section presents the results according to the two research questions of this study: a) the relationship between appropriateness perception and listeners' L1 and b) the relationship between appropriateness perception and speakers' phonological and pragmatic features. The first research question investigated the extent to which differences were observed in the appropriateness ratings by listeners of different L1s. A linear mixed-effects model was computed, and the overall model explained 48.67% of the variance (conditional R^2), with listeners' L1 explaining 5.52% of the variance (marginal R^2).

After the random effects had been controlled for, the estimated marginal means of listener ratings of appropriateness were computed (see Table 2). Based on the model results, it turned out that L1 Spanish and L1 English listeners tended to provide more lenient ratings, compared to L1 listeners of Chinese and Indian languages. Notably, the confidence intervals of these estimated marginal means were quite large. To confirm any statistical differences, a pairwise comparison adjusted for family-wise error rate was computed, and results did not show a statistically significant difference across listeners of any L1.

Table 2 Estimated marginal means of listener appropriateness ratings controlling for random effects

Listener L1	EMMeans	SE	95% CI
Indian languages	4.62	0.76	[3.14, 6.10]
Chinese	4.76	0.58	[3.63, 5.90]
English	5.53	0.58	[4.39, 6.67]
Spanish	5.63	0.60	[4.46, 6.80]

5.2. Appropriateness and phonological and pragmatic features

The second research question explored the relationship between appropriateness and phonological and pragmatic features. A linear mixed-effects model was computed, and the overall model explained 55.62% of the variance (conditional R^2), with listeners' L1 explaining 14.07% of the variance (marginal R^2).

Of the fixed effects, space (i.e., proportion of stressed words) and pragmatic strategies contributed significantly to listeners' appropriateness ratings. Specifically, space was negatively correlated with the ratings. In terms of the pragmatic strategies used, the estimated marginal means of listener ratings of appropriateness were computed given its nine levels (see Table 3). Based on the estimated marginal means, it seems that two pragmatic strategies stood out because they were associated with higher appropriateness scores: a) expressing wish ($EMMean = 6.42$, 95% CI [5.58, 7.26]) and b) statement of regret/apology ($EMMean = 5.84$, 95% CI [5.00, 6.67]). Other pragmatic strategies seemed to be receiving relatively homogenous ratings ($EMMean = 4.74$ - 5.36). A pairwise comparison was not computed because this categorical variable entailed nine levels, and adjustment for family-wise error rates would be too stringent with too many levels (see e.g., Groenwold et al., 2021).

Table 3 Estimated marginal means of appropriateness ratings per different pragmatic strategies

Pragmatic Strategies	EMMeans	SE	95% CI
expressing wish	6.42	0.43	[5.58, 7.26]
statement of regret/apology	5.84	0.43	[5.00, 6.67]
expression of disagreement	5.36	0.43	[4.52, 6.20]
provision of alternatives	5.26	0.42	[4.44, 6.09]
use of plain indirect statements	5.19	0.41	[4.39, 5.99]
statement of principle/philosophy	5.07	0.41	[4.25, 5.88]
provision of reason	5.04	0.40	[4.25, 5.83]
use of non-performative statements/negotiation of proposition	4.98	0.41	[4.19, 5.78]
bluntness/use of performative verbs	4.74	0.46	[3.85, 5.64]
provision of advice	n/a		
avoidance	n/a		

Note. The coding suggested no use of: a) the provision of advice or b) the avoidance strategies

After the fixed effects had been controlled for, the contribution of random effects to the appropriateness ratings was computed. Listener idiosyncrasies accounted for the most variance explained (26.07%), followed by speaker idiosyncrasies (7.93%), listener L1 (7.64%), and speaker gender (5.21%). Other factors, such as speaker L1 (0.03%), listener gender (0.06%), and the pragmatic scenario (1.02%), did not seem to contribute to the appropriateness ratings significantly.

6. Discussion

6.1. Perception of appropriateness and listener L1 background (RQ 1)

This section seeks to interpret the results observed in the present study in light of the previous research findings. This discussion is organized based on the two research questions of the study: a) appropriateness perception in relation to listener L1 background and b) appropriateness perception in relation to speakers' phonological and pragmatic features. First, findings suggest that listeners' L1 background did not affect their assigned appropriateness significantly. Moreover, the 95% confidence intervals around the estimated marginal means seemed quite large (Table 2). To interpret the findings, it is possible that beyond listener L1, other features could potentially impact the appropriateness ratings more. From the speakers' perspective, this can be explained by the prosodic and pragmatic features presented in the utterances (as was suggested in the following findings). From the listeners' perspective, it is possible that besides listener L1, other variables can come into play and influence their speech perception, for example, listeners' accent familiarity (Miao, 2023), multilingual experience (Saito & Shintani, 2015), and linguistic training (Saito et al., 2017). Future research could further explore how listener-based variables could influence their appropriateness perception, which this study does not focus on.

6.2. Perception of appropriateness and prosodic and pragmatic features (RQ 2)

Space was negatively correlated with listener appropriateness ratings. That is, the more words that were stressed in a refusal, the lower the ratings tended to be. This finding corroborated previous research based on theoretical and anecdotal evidence suggesting a link between prosody and pragmatics (see e.g., Brazil, 1997; Hirschberg, 2017). There are several possible explanations for this finding. Over-stressing is a common phenomenon of L2 speech, especially in speakers whose L1 is syllable-timed, not stress-timed (see Pickering, 2001; Wennerstrom,

1998), and rhythm was found to be highly correlated with L2 speakers' comprehensibility (see Hahn, 2004; Isaacs & Trofimovich, 2012). Thus, it is possible that over-stressing reduced the speaker's comprehensibility leading to lower appropriateness ratings. Although this speculation needs future empirical support, it tentatively reveals that comprehensibility might be a building block of appropriateness (see Hymes, 1971): without knowing what the speaker said, it is difficult to assess how appropriately they said it for a certain context. Another explanation was that speakers' over-stressing words in refusals elicited listeners' stereotypes associated with this linguistic feature (see Lindemann et al., 2014), which could explain the lower ratings assigned. Nonetheless, this finding should be treated with caution because we speculate that the relationship between space and appropriateness and/or comprehensibility may not be linear. That is, whether an L2 speaker does not stress the right content words in a request or over-stresses words (i.e., emphasizes all words including function words), it may negatively affect comprehensibility. It is likely that we only captured one part of the picture and further investigation of the relationship between comprehensibility and appropriateness is needed.

Tone choice and pitch range did not seem to have influenced listener appropriateness ratings when other variables were controlled for. The coding procedure might explain this difference. In the present report, tone unit was only coded in the main strategy within the speech act. However, listeners were exposed to other expressions surrounding the main strategy, including mitigating devices. Had the research team also coded for the overall tone choice as another variable, a different picture may have emerged. Comparatively, although an overall pitch range of requests was calculated, it is possible that in perceiving the refusals, listeners also attended to the pitch range of prominent syllables specifically (ones that carried more meaning). Therefore, it may be beneficial to conduct additional analyses to include: a) tone choice of the core sentence of the speech act, b) overall tone choice throughout the stimuli in percentages, c) pitch range based on prominent syllables, and c) overall pitch range. This change may provide a more comprehensive picture of listener perception of appropriateness and lead to a more thorough discussion.

Pragmatic strategies in refusals, especially: a) expressing wish and b) expressing regret/apology, seemed to be correlated with listener appropriateness ratings. To explain this finding, when speakers expressed "wish," they exclusively relied on the formulaic sequence "I wish I could" in our data. This sequence may be judged as especially appropriate by listeners (see Taguchi, 2013). Moreover, refusals are face-threatening in nature, and this was magnified in the target pragmatic scenario where speakers were actually talking to a higher-power interlocutor (i.e., professor and supervisor). Thus, making apologies might have

been perceived as especially useful in talking to higher-status persons, hence more preferred in the appropriateness evaluation. This conclusion needs future empirical support, and it also reveals the potential mediating effect of pragmatic situations on the appropriateness-strategies relationship.

In terms of listener idiosyncrasies, we found that ratings of appropriateness fluctuated from listener to listener. To explain this, we argue that the construct of appropriateness, as with comprehensibility and intelligibility, is co-constructed by both speakers and listeners (see Rajadurai, 2007). In other words, there is a symbiotic relationship between speakers and listeners in the construction and conceptualization of appropriateness, and it would be less ideal to assume that being able to communicate appropriately is solely a speaker's responsibility. This echoes and provides empirical support to recent arguments stating that L2 English learners should not be viewed as less competent or appropriate whilst basing the criterion against the L1-speaking, high-power communities (see Flores & Rosa, 2015, forthcoming).

Finally, we found it interesting that speakers' L1 status did not seem to influence the appropriateness ratings. This finding seems to contrast with previous findings suggesting that oftentimes L2 speakers were negatively evaluated on a range of measures from their linguistic competence and suitability of employment (see Ghanem & Kang, 2021; Kang & Rubin, 2009). It is possible that the variance that speaker L1 status could have explained was overshadowed by the rhythm measure in the model (i.e., space). In other words, space might have already explained the variances that speakers' L1 could explain. Another possible explanation is that in assessing appropriateness, listeners did not just attend to phonological traits such as accents and perceived comprehensibility. Instead, they additionally focused on the intended meaning when assessing appropriateness. If this is the case, then this again supports the argument that both "what is said" and "how it is said" are important in pragmatic comprehension (see Hymes, 1971). However, it is important to note that the L2 speakers included in the study were highly intelligible; therefore, it is likely that the presence of L2 traits was less pronounced and thus did not significantly affect the ratings.

6.3. Pedagogical implications

Based on the findings obtained in the present study, we encourage L2 English educators and material developers to design instructional materials that incorporate guidelines for learners to make speech acts more appropriate. To do this, students should be aware that it is not just what is said that is important, but also how it is said. In line with this reasoning, teachers could draw students'

attention to the idea that pragmatic success (i.e., saying things appropriately given the situational context and interlocutors) depends on many different factors, including the use of both phonological features and pragmatic features.

On the phonological level, “space” was found to be a significant predictor of listener perception of appropriateness. This means that learners’ choice of where to put emphasis in speech acts could influence the perceived appropriateness of their utterances. It is thus important to provide learners with explicit instruction and meaningful perception- and production-based exercises to stress syllables in speech acts strategically (e.g., stressing content words that are important for interlocutors’ understanding of a request or a refusal). Beyond speech acts, pragmatic functions of prosody need to be consistently included as a part of spoken interaction practice. On the lexico-grammatical level, some pragmatic strategies contributed to higher appropriateness perception scores, including formulaic sequences expressing wishes (e.g., *I wish I could . . .*) and apologies (e.g., *I am sorry but . . .*) when making a refusal. Taken together, instructors could introduce these pragmatic strategies to help learners produce more appropriate speech acts, leading to more successful interactional experience. In our view, it is vital to incorporate the prosodic component into instruction about pragmatic strategies. Nonetheless, given the number of studies exploring factors affecting listeners’ pragmatic perception is still relatively low, these pedagogical implications above should be taken with caution.

6.4. Limitations and future directions

The findings in this study need to be considered with certain limitations in mind. First, the speaker sample only included L1 English and L1 Chinese. Thus, the findings may not be applicable beyond these language varieties. Future research including more spoken varieties of English is welcome to extend the findings observed in the current study. Second, the study only included two refusal scenarios (i.e., student-professor and employee-supervisor). Future studies could include different speech acts (e.g., requests) to improve the generalizability of the findings. Last, the listener sample only included a limited number of English varieties, with an especially small sample of L1 Indian languages. Future studies might build on the current study and explore a more diverse listener sample in the perception of appropriateness.

7. Conclusion

Two overarching findings were observed in this study. First, speakers’ phonological/prosodic features indeed made a unique contribution to listeners’ perception

of appropriateness in refusals. This provides empirical evidence, corroborating many theoretical arguments, that pragmatics and prosody are two related sub-fields. Future research investigating this interface would be especially fruitful, given the scarcity of current relevant empirical research. Second, pragmatic strategies, especially apologies and expressing wishes, considerably contributed to the perception of appropriateness in refusals. This, coupled with the first finding, revealed that appropriateness might have been more complex than previously assumed, with different variables feeding into it. It is likely that this construct is multi-dimensional with different sub-constructs (e.g., phonology, pragmatic strategies/lexicogrammar). Future research might want to build on the current study, exploring factors affecting appropriateness in a more holistic fashion. This would offer theoretical implications regarding the conceptualization of appropriateness.

The findings of the study yield several practical implications including specifically the importance of prosody instruction in language teaching and learning. At the moment, pronunciation teaching seems to be heavily focused on segmental instruction (e.g., vowels and consonants). This paper, along with many others, provides empirical insights supporting a more theory-oriented pronunciation instruction with an emphasis of suprasegmental features (e.g., rhythm). Second, and in a similar vein, language teaching and learning should contextualize prosodic aspects of interaction within pragmatic instruction. Pragma-prosodic interface has been largely overlooked in classroom teaching but has an important role in communication in real-life settings. Overall, we hope the present research has not only shed light on the prosody-pragmatics link, but will also encourage more research to build on it at conceptual and pedagogical levels.

References

- Alemi, M., Eslami, Z. R., & Rezanejad, A. (2014). Rating EFL learners' interlanguage pragmatic competence by non-native English-speaking teachers. *Procedia – Social and Behavioral Sciences*, 98, 171-174. <https://doi.org/10.1016/j.sbspro.2014.03.403>
- Alemi, M., & Khanlarzadeh, N. (2016). Pragmatic assessment of request speech act of Iranian EFL learners by non-native English-speaking teachers. *Iranian Journal of Language Teaching Research*, 4(2), 19-34.
- Alemi, M., & Motamedi, M. (2019). Pragmatic criteria in the holistic and analytic rating of the disagreement speech act of Iranian EFL learners by non-native English speaking teachers. *TESOL Quarterly*, 38(1), 1-36. <https://doi.org/10.22099/jtts.2020.34820.2738>
- Al Masaeed, K., Taguchi, N., & Tamimi, M. (2020). Proficiency effects on L2 Arabic refusals: Appropriateness, linguistic strategies and multidialectal practices. *Applied Pragmatics*, 2(1), 26-53. <https://doi.org/10.1075/ap.19007.mas>
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Beebe, L.M., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. C. Scarcella, E. S. Anderson, & S. D. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 55-94). Newbury House.
- Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* [Computer program]. Version 6.0. 37.
- Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge University Press.
- Cunningham, D. J. (2017). Second language pragmatic appropriateness in telecollaboration: The influence of discourse management and grammaticality. *System*, 64, 46-57. <https://doi.org/10.1016/j.system.2016.12.006>
- Dewaele, J.-M. (2008). "Appropriateness" in foreign language acquisition and use: Some theoretical, methodological and ethical considerations. *International Review of Applied Linguistics in Language Teaching*, 46(3), 235-255. <https://doi.org/10.1515/IRAL.2008.011>
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2022). *Ethnologue: Languages of the world*. SIL International.
- Félix-Brasdefer, J. C. (2003) Declining an invitation: A cross-cultural study of pragmatic strategies in American English and Latin American Spanish. *Multilingua*, 22, 225-255. <https://doi.org/10.1515/mult.2003.012>
- Flores, N., & Rosa, J. (forthcoming). Undoing competence: Coloniality, homogeneity, and the overrepresentation of whiteness in applied linguistics. *Language Learning*. <https://doi.org/10.1111/lang.12528>

- Gass, S., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34(1), 65-87. <https://doi.org/10.1111/j.1467-1770.1984.tb00996.x>
- Ghanem, R., & Kang, O. (2021). Groenwold, R. H. H., Palmer, T. M., & Tilling, K. (2021). To adjust or not to adjust? When a "confounder" is only measured after exposure. *Epidemiology*, 32(2), 194-201. <https://doi.org/10.1097/EDE.0000000000001312>
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201-223. <https://doi.org/10.2307/3588378>
- Herrero, C., & Devis, E. (2020). Unintentional impolite intonation in L2 Spanish requests produced by Chinese workers living in Madrid. *Proceedings of the 10th International Conference on Speech Prosody 2020*, Tokyo, Japan, 848-852.
- Hirschberg, J. (2017). Pragmatics and prosody. In Y. Huang (Ed.), *The Oxford handbook of pragmatics* (pp. 1-19). Oxford University Press.
- Hymes, D. (1971). *On communicative competence*. University of Pennsylvania Press.
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475-505. <https://doi.org/10.1017/s0272263112000150>
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301-315. <https://doi.org/10.1016/j.system.2010.01.005>
- Kang, O., & Rubin, D. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28, 441-456. <https://doi.org/10.1177/0261927X09341950>
- Kang, O., Yaw, K., & Kostromitina, M. (2023). The effects of situational contexts and occupational roles on listeners' judgements on accented speech. *Psychology of Language and Communication*, 27(1) 1-22. <https://doi.org/10.58734/plc-2023-0001>
- Kostromitina, M. (in press). An exploratory study of intonation variation in L1 and L2 English speakers' pragmatic production of high imposition requests and refusals. *Applied Pragmatics*.
- Krulatz, & Dixon, T. (2020). The use of refusal strategies in interlanguage speech act performance of Korean and Norwegian users of English. *Studies in Second Language Learning and Teaching*, 10(4), 751-777. <https://doi.org/10.14746/ssllt.2020.10.4.5>
- Lindemann, S., Litzenberg, J., & Subtirelu, N. (2014). Problematizing the dependence on L1 norms in pronunciation teaching: Attitudes toward second-language accents. In J. Levis & A. Moyer (Eds.), *Social dynamics in second*

- language accent (pp. 171-194). De Gruyter Mouton. <https://doi.org/10.1515/9781614511762.171>
- Miao, Y. (2023). The relationship among accent familiarity, shared L1, and comprehensibility: A path analysis perspective. *Language Testing*, 40(3), 723-747. <https://doi.org/10.1177/02655322231156105>
- Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, 35, 233-255. <https://doi.org/10.2307/3587647>
- Pickering, L. (2018) Discourse intonation: A discourse-pragmatic approach to teaching the pronunciation of English. *Journal of Second Language Pronunciation*, 5(2), 328-332. <https://doi.org/10.1075/jslp.00012.bol>
- Prieto, P., & Rigau, G. (2012). Prosody and pragmatics. In L. Payrató & J. Cots (Eds.), *The pragmatics of Catalan* (pp. 17-48). De Gruyter Mouton. <https://doi.org/10.1515/9783110238693.17>
- Purpura, J. (2004). *Assessing grammar*. Cambridge University Press.
- Rajadurai, J. (2007). Intelligibility studies: A consideration of empirical and ideological issues. *World Englishes*, 26(1), 87-98. <https://doi.org/10.1111/j.1467-971X.2007.00490.x>
- Saito, K., & Shintani, N. (2015). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly*, 50(2), 421-446. <https://doi.org/10.1002/tesq.234>
- Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2017). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141-156). Multilingual Matters. <https://doi.org/10.21832/ISAACS6848>
- Salazar Campillo, P., Safont-Jordà, P. M., & Codina Espurz, V. (2009). Refusal strategies: A proposal from a sociopragmatic approach. *Revista Electrónica de Lingüística Aplicada*, 8, 139-150.
- Sydorenko, T., Maynard, C., & Guntly, E. (2014). Rater behaviour when judging language learners' pragmatic appropriateness in extended discourse. *TESL Canada Journal*, 32(1), 19-41. <https://doi.org/10.18806/tesl.v32i1.1197>
- Taguchi, N. (2002). An application of relevance theory to the analysis of L2 interpretation Processes: The comprehension of indirect replies. *IRAL – International Review of Applied Linguistics*, 40(2), 151-176. <https://doi.org/10.1515/iral.2002.006>
- Taguchi, N. (2006). Analysis of appropriateness in a speech act of request in L2 English. *Pragmatics*, 16(4), 513-533. <https://doi.org/10.1075/prag.16.4.05tag>
- Taguchi, N. (2011). Rater variation in the assessment of speech acts. *Pragmatics*, 21(3), 453-471. <https://doi.org/10.1075/prag.21.3.08tag>

- Taguchi, N. (2012). *Context, individual differences and pragmatic competence*. Multilingual Matters.
- Taguchi, N. (2013). Refusals in L2 English: Proficiency effects on appropriateness and fluency. In O. Marti-Arnandiz & P. Salazar-Campillo (Eds.), *Refusals in instructional contexts and beyond* (pp. 101-119). Brill.
- Taguchi, N., Hirschi, K., & Kang, O. (2022). Longitudinal L2 development in the prosodic marking of pragmatic meaning: Prosodic changes in L2 speech acts and individual factors. *Studies in Second Language Acquisition*, 44(3), 843-858. <https://doi.org/10.1017/S0272263121000486>
- Tajeddin, Z., & Alemi, M. (2014). Pragmatic rater training: Does it affect non-native L2 teachers' rating accuracy and bias? *Iranian Journal of Language Testing*, 4(1), 66-83.
- van Compernelle, R. A. (2014). *Sociocultural theory and L2 instructional pragmatics*. Multilingual Matters.
- Vanderplank, R., (1993). Pacing and spacing as predictors of difficulty in speaking and understanding English. *ELT Journal*, 47, 117-125. <https://doi.org/10.1093/elt/47.2.117>
- Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, 42, 1-13. <http://www.jstor.org/stable/44486381>
- Wharton, T. (2012). Pragmatics and prosody. In K. Allan & K. M. Jaszczolt (Eds.), *The Cambridge handbook of pragmatics* (pp. 567-584). Cambridge University Press. <https://doi.org/10.1017/CBO9781139022453.031>