Open Library of Humanities

Prosodic realization and interpretation of English imperatives

Elise McClay, Department of Linguistics, University of British Columbia, CA, emcclay@sfu.ca

Megan Keough, Department of Linguistics, University of British Columbia, CA, megan.e.keough@gmail.com

Molly Babel, Department of Linguistics, University of British Columbia, CA, molly.babel@ubc.ca

Lisa Matthewson, Department of Linguistics, University of British Columbia, CA, lisa.matthewson@ubc.ca

Imperative clauses can communicate a number of speech acts, and differences in intonation have been argued to prompt different interpretations. So far, however, limited phonetic evidence has been presented for such proposals. The focus in the current work is on maximally strong imperatives (commands) and weaker imperatives (mainly involving advice) in English. We report on a series of phonetic experiments intended to address whether listeners reliably associate stronger and weaker imperatives with idealized intonation (Experiment 1) and whether speakers produce these two types of imperatives differently (Experiment 2). Individual variability in our production data led us to test whether listeners can map the variable pronunciation patterns found in Experiment 2 to stronger and weaker imperatives (Experiment 3), as they did with the idealized pronunciations in Experiment 1. Despite substantial cross-talker intonation variation, listeners' stronger/weaker imperative recognition performance paralleled accuracy with the idealized productions. Analysis of the whole utterances indicates that speech rate and global pitch setting work along with the final intonation contour to signal semantic meaning in English imperatives. These results suggest that the mapping between meaning and form is complex and involves redundancy.

1 Introduction

English imperative clauses can be used to convey a range of speech acts, including commanding, requesting, warning, giving advice, and giving permission. Some of these different functions are illustrated in (1).

(1)   a.   Read this!   COMMAND
      b.   Turn on the light, please.   REQUEST
      c.   Take the A train if you want to go to Harlem.   ADVICE
      d.   (It starts at eight, but) come earlier if you like!   PERMISSION

                                             (adapted from Kaufmann 2012: 12–13)

Although the lexical content of an imperative can influence the most likely speech act it performs, the same string can also be used to perform a variety of different speech acts, as shown for the sentence in (2) in the contexts in (2a–d).

(2)   Turn on the light.
      a.   From a sergeant to a corporal.   COMMAND
      b.   Asking a friend who is closer to the light switch.   REQUEST
      c.   Offering a tip about how to see better to read.   ADVICE
      d.   I'm not sleeping, so go ahead if you want to, …   PERMISSION

Some researchers have proposed that intonation can be used to disambiguate the different interpretations of imperatives; see for example Chatzikonstantinou (2013), Oikonomou (2016), Jeong & Condoravdi (2017), Portner (2018), and Rudin (2018). With the exception of Oikonomou's work on Greek, all the prior studies focus on the intonation/meaning connection in English imperatives. Claims have been made that command imperatives are prosodically distinct from permission imperatives (Oikonomou 2016) and even from advice imperatives (Chatzikonstantinou 2013, and some indications in Oikonomou 2022). For Portner (2018), the intonational difference is claimed to correlate with whether the content of the imperative is presented as being the speaker's priority (as with commands), or the addressee's (as with permission, invitations, and some types of advice). However, no full studies have been carried out to determine the exact intonational properties of imperatives, nor to establish which intonation patterns associate with which semantic or pragmatic properties.

In this paper we report on a series of phonetic experiments designed to address the questions in (3). The questions are in principle cross-linguistic and need to be answered for a diverse range of languages, but here we focus on English.[1]

---

[1] We generically refer to 'English' because while our study was conducted in western Canada where Canadian English is the local standard, our requirement for participants is that they self-identify as a native speaker of any variety of English. This comes with the assumption that they are familiar with the local variety of English on which we base our materials.

(3) i.   Do speakers produce 'command' and 'non-command' imperatives with different prosody?
     ii. If yes, what are the prosodic characteristics of the different uses of imperatives?
     iii. Do listeners use these production cues in their perception in order to disambiguate between different types of imperative?
     iv. How should prosody-semantics interactions in imperatives be analyzed?

Our studies suggest that English speakers do produce the two types of imperatives with different prosody, and that English listeners are able to use production cues in their perception in order to disambiguate meaning. However, we found that there is substantial variability in the intonational contours used by speakers, making it difficult to assign particular meanings to specific intonational melodies. Instead, we discovered some robust patterns, which we refer to as "macro" prosodic settings, involving overall speech rate and pitch height, that correlate with the command/advice imperative distinction. An examination of productions that elicit unanimous agreement in their semantic meaning also suggests that the realization of the nuclear pitch accent in imperatives cues the strong or weak reading. This leads us tentatively to suggest that there are prosodic morphemes that consist not solely of particular tonal contours, but instead of macro settings involving overall speech rate and pitch height and nuclear pitch accent.

In section 2, we provide some brief semantic and prosodic background, and in section 3 we give an overview of prior contributions on the prosody-meaning connection in imperatives. Sections 4 through 6 report on our three experiments. Experiment 1 investigated whether listeners associate 'idealized' distinct prosodies with command and non-command imperatives (mostly advice uses, with some permission tokens). Experiment 2 was a production experiment, eliciting imperatives from participants. Experiment 3 tested an independent group of listeners' interpretation of the natural productions from Experiment 2. In Section 7 we examine the prosodic patterns of a subset of utterances. We conclude in Section 8 with speculations about the implications of our results for the analysis of prosody-meaning interactions.

2 Semantic and prosodic background

The optimal analysis of the semantics and pragmatics of imperatives is a matter of lively debate in current literature. One major point of contention concerns whether imperatives semantically contain modal operators (Han 1999; 2019; Schwager 2006; Grosz 2009; Condoravdi & Lauer 2012; Kaufmann 2012; Kaufmann & Kaufmann 2016; Oikonomou 2016; 2022; among others), or whether their directive force comes from the pragmatics instead (Portner 2004; 2007; 2018; Charlow 2010; 2011; Starr 2010; 2020; Roberts 2015; von Fintel & Iatridou 2017, among others). A typical modal analysis can be (very roughly and informally) paraphrased as in (4), while a pragmatic analysis might be (similarly roughly and informally) summarized as in (5).

(4)   Modal approach:
       An imperative containing the property P truth-conditionally conveys 'You must/should
       P' (and in addition presupposes that the context is such that an utterance of 'You must/
       should P' is performative, i.e., functions to impose the obligation).

(5)   Pragmatic approach:
       An imperative expressing the property P updates the discourse context so that P is
       added to the addressee's set of future commitments or preferences.

An important issue, faced by both strands of analysis, is how imperative constructions can receive a unified analysis, given their wide range of uses in discourse and in particular the difference between 'strong' readings — like commands, requests, and some types of advice — and 'weak' readings — like disinterested advice, invitations, and permission. Within a modal analysis, for example, the issue manifests itself as the fact that permission imperatives (like in (2d)) are not accurately paraphrased with the necessity modal *must,* and seem at first glance to reflect instead a possibility modal like *may*.

The majority response to this problem has traditionally been to analyze the necessity readings as basic, and to derive permission (possibility) readings through a pragmatic weakening mechanism (e.g., Schwager 2006/Kaufmann 2012). Recently, some authors have gone in the other direction, proposing a basic interpretation which allows permission readings, plus a pragmatic strengthening mechanism to obtain necessity readings (Oikonomou 2016; 2022; Francis 2020). A third option is that there is a real ambiguity, so that necessity and possibility imperatives actually have different denotations; see e.g., Grosz (2009) and Carter (2021) for proposals of this type.

Turning to the prosody-meaning connection, many authors have proposed the existence of prosodic elements whose presence changes the semantics or pragmatics of an utterance. See, for example, Pierrehumbert & Hirschberg (1990); Gussenhoven (2004); Truckenbrodt (2012, 2020); Bartels (2014), among many others. Some of the semantic or pragmatic phenomena that have been claimed to be influenced by prosody include quantifier scope (e.g., Jackendoff 1972; Surányi & Turi 2020), rising declaratives (e.g., Gunlogson 2001, 2008; Fletcher & Loakes 2010; Ritchart & Arvaniti 2013; Heim 2019; Bhadra 2020), information structural notions (Steedman 1991; 2007; 2014), and in general, the management of interlocutors' commitment and/or belief states (e.g., Davis 2009; Malamud & Stephenson 2015; Rudin 2018; Heim 2019; Bhadra 2020).

These and similar phenomena have led to the hypothesis that there are intonation-meaning mappings (i.e., prosodic morphemes) that contribute specific meanings apart from the sentence's segmental material (Pierrehumbert & Hirschberg 1990; Steedman 1991; Gussenhoven 2004; i.a.). If such prosodic morphemes exist, one possibility is that stronger and weaker imperatives

in English are distinguished by a prosodic difference (for instance, H% vs. L%),[2] but before we make this conclusion, we need evidence of systematicity between imperative prosody and interpretation. In the following section we provide an overview of prior work in the area of English imperative intonation.

3 The prosody-meaning connection in imperatives

In this section we review prior proposals about the correlation between interpretation and intonation in imperatives. We begin with the paper that inspired the current research, Portner (2018).

Portner's (2018) approach to imperatives, building on his earlier work (2004; 2007), involves a structured model of the context that captures interlocutors' commitments to both factual information and priorities. Mutual commitments to factual information are represented in the Common Ground, and mutual commitments to future priorities of interlocutors are provided by the mutual To-Do List. Each interlocutor also has their own individual commitments, both to facts and to priorities. Portner's (2018) novel contribution is to propose that there is a difference in conventional effect between 'falling imperatives' and 'rising imperatives', parallel to the falling/ rising declarative distinction (Gunlogson 2001, among many others). Both falling and rising imperatives propose an addition to the addressee's To-Do List, but in addition, 'Rising imperatives propose the *addressee's* commitment to treating the imperative's content as a priority, while falling imperatives propose the *speaker's* commitment' (Portner 2018: 307; emphasis added).

To illustrate, Portner's proposal is that the imperative *Have a seat,* when uttered with falling intonation, is used when the speaker is directing the addressee to sit down, and is not taking the addressee's priorities into account. When uttered with rising intonation, the same string *Have a seat* invites or gives permission to the addressee to sit down, under the speaker's assumption that sitting down will benefit the addressee. In the falling case, 'the *speaker* rates futures in which the addressee sits down higher than those in which he does not,' and in the rising case, 'the *addressee* rates futures in which the addressee sits down higher than those in which he does not' (Portner 2018: 308; emphasis added). In both cases, the speaker expects that the priority for the addressee to sit down will become mutual.

---

[2] In this paper we make use of the ToBI (Tones and Break Indices) system for annotating prosodic contours (Silverman et al. 1992). The ToBI symbols relevant to the current research are listed in (6), where H stands for "high" and L for "low".

(6) H*, L*: pitch accents (on the nuclear stress)
    H-, L-: phrase accent (between pitch accents and edges of intonation phrases)
    H%, L%: boundary tone at the right edge of intonation phrase

A typical English falling contour is annotated as H* L- L%, and a typical rising contour is L* H- H%.

In terms of the speech acts introduced at the beginning of this paper, commands fall into the falling group, while invitations or offers like *Have a cookie!* fall into the rising group. Although Portner does not specifically discuss advice, his characterization of rising imperatives as offering a priority that the speaker believes will benefit the addressee suggests that at least some cases of advice will be expected to have rising intonation.[3]

While Portner himself did not conduct any phonetic studies of imperative intonation, a few other authors have. Chatzikonstantinou (2013) conducted three pilot studies of English imperatives, testing the hypothesis that there is an intonational difference between imperatives used to convey commands and those used to give advice. He proposes that commands and advice are associated with distinct pitch contours during production, and that "during perception listeners can consistently identify whether an imperative is a command or an advice based on the acoustic signal" (Chatzikonstantinou 2013: 4).

In terms of production, Chatzikonstantinou found that command imperatives were uttered with "a flat intonation following H H L- (L%) pattern" (2013:6), while advice imperatives received "a H H*L- (H%) with a distinctive large excursion at the N[oun]" (2013: 7). It should be noted however that these results are based on analysis of the data of just one (female) speaker. Chatzikonstantinou's two pilot perception studies relied on the edited productions of that speaker, and specifically used the final portions of imperatives embedded within larger utterances (e.g., She told him to read the book). Chatzikonstantinou found that participants were able to correctly identify whether an imperative was a command or advice with at least 78.5% accuracy.

Oikonomou (2016) conducted a small production experiment to test the hypothesis that command and request interpretations of imperatives involve broad focus intonation, and permission imperatives do not. Five English speakers and four Greek speakers read aloud 10 pairs of imperative sentences; within each pair, the strings were identical but one context supported a command/request reading while the other supported a permission reading. The "prevailing pattern" in the English productions was that command/request imperatives were realized with a high Nuclear Pitch Accent followed by a low boundary tone ((H*) H* L-L%); this parallels the intonation of declaratives with broad focus (2016: 70–71). In "a few cases", English command/request imperatives involved rising intonation (2016: 71). Similarly in Greek, the predominant pattern was for command/request imperatives to end in a H* L-L% contour, with a "second prevailing pattern" involving a final rising intonation (Oikonomou 2016: 72–73).

Oikonomou further found that permission imperatives in both English and Greek were typically realized with a high Nuclear Pitch Accent on the verb (i.e., earlier in the imperative than for commands/requests), followed by de-accenting (2016: 74). Based on the information

---

[3] Portner analyzes disinterested advice and 'for example' imperatives as a separate phenomenon: they 'do not seek to update the addressee's to-do list' (Portner 2018: 310).

given, it therefore seems that the difference between the two types was not one of contour, but of the placement of the H*.

Oikonomou's semantic explanation for why these intonation patterns should correlate with these readings relates to a claim that the two types of imperatives involve focus on different constituents; see her discussion for details. With respect to the rising intonation found on some productions of command/request imperatives, she argues that "rising intonation is a marked prosodic pattern which can apply equivalently to all types of Imperatives to encode uncertainty/ level of endorsement etc. but it doesn't differentiate the permissions from the commands" (2016: 66). This contrasts directly with Portner's proposal that rising intonation *does* signal permission uses.

Rudin (2018: 20) argues that falling intonation (more precisely "steeply, monotonically falling intonational tunes", or H* L-L%) signals that the speaker is making a commitment by virtue of their utterance, while rising intonation ("steeply, monotonically rising intonational tunes", L* H-H%) conveys that the speaker is not making a commitment (see also Pierrehumbert & Hirschberg 1990; Truckenbrodt 2012). Within imperatives, falling intonation conveys that the speaker is committed to the claim that the addressee should perform the relevant action; rising intonation signals that the speaker is leaving it up to the addressee to decide whether or not they should perform the action. Again, this differs from Portner's characterization of the semantic division. Unlike for Portner, for Rudin the falling/rising split does not correlate with a speech act division. Rudin also argues against Portner's claim that a rising imperative signals that the speaker thinks the addressee wants to perform the action.

In support of the uncoupling of intonation from speech acts, Rudin offers data that suggest that commands and offers with falling intonation pattern together, in opposition to offers with rising intonation. For example, offers with falling intonation can be followed by *I insist* (7), but offers with rising intonation cannot (8).

(7)   A: Have a cookie.   OFFER WITH FALLING INTONATION
        B: No, thanks.
        A: I insist.   (Rudin 2018: 97)

(8)   A: Have a cookie?   OFFER WITH RISING INTONATION
        B: No, thanks.
        A: ?? I insist.   (Rudin 2018: 97)

Jeong & Condoravdi (2017) carried out a full perception experiment to test a hypothesis about the intonation of English imperatives in various types of discourse contexts. (See also Jeong & Condoravdi 2018.) Specifically, they investigated the 'downstepped level terminal contour' (DLT), H* !H-L%. This is also known as the 'calling contour', and is often found in vocatives

(*Jacob! Your lunch!*). In the experiment, 400 participants heard imperatives which had been recorded by native speakers and then synthetically manipulated to have the desired intonational properties (either the DLT contour, or the more traditional H* L-L% or L* L-L%). Participants judged which recordings fit best into scenarios they read.

Jeong & Condoravdi found that the DLT contour is usually compatible with imperatives expressing well-wishes (e.g., *Enjoy the movie!*), but incompatible with orders and offers. However, they also argue that the appropriateness of the DLT contour crosscuts speech-act types and is systematically context-dependent. In particular, DLT is used when the speaker wishes to signal that she does not expect her own future actions to affect whether the content of the imperative clause is realized. This accounts for the finding that requests or warnings which are reminders license DLT, but novel requests or warnings about unfamiliar dangers typically do not (cf. also Ladd 1978 for this empirical generalization).

While Jeong & Condoravdi's work is, to our knowledge, the most robustly phonetically grounded prior work on imperatives and intonation, we set it aside from now on because it deals with a particular contour that does not (as they convincingly show) correlate with the command vs. non-command distinction which is our focus here.[4]

Finally, a very recent pilot study by Oikonomou (2022) has detected a potential three-way intonational contrast in Greek imperatives. In the production portion of the experiment, a single speaker was asked to read imperative utterances as naturally as she could, in both command and permission contexts. Next, 23 participants in a perception study were asked to categorize each imperative production as either a 'requirement' or a 'possibility'. Oikonomou reports that one intonational pattern was produced for the permission imperatives (a Nuclear Pitch Accent on the verb), but two different patterns were produced for the command contexts: one ending with a final L-L% boundary tone, and one ending with a final L-H% boundary tone. Participants in the perception study appear able to distinguish the commands from the permissions, and the commands ending in L-H% were judged as slightly more 'requirement'-like than the L-L% commands.[5]

---

[4] Other phonetically grounded research on the prosody-meaning connection includes Jeong & Potts (2016), who conducted perception experiments to probe how English-speaking listeners interpreted declaratives and interrogatives (plus a few imperatives), with three different prosodic contours: falling, level and rising. They also asked participants how annoyed, authoritative, and polite the speakers sounded. They found that across all sentence-types, falling tunes led to the speaker being interpreted as authoritative, level tunes correlated with sounding annoyed, and rising tunes were perceived as polite. Petrone & D'Imperio (2011) also found that a falling tune H*L-L% on request/offer utterances (e.g., Can [you/I] check the weather for [me/you]?) was perceived as increasing speaker authority, particularly for requests. We briefly return to the interaction of speaker authority with phonetics in section 8. However, as our results will show melody differences only in the nuclear pitch accent, we cannot utilize these proposals from Petrone & D'Imperio and Jeong & Potts.

[5] No statistical model is described or reported in Oikonomou (2022).

Summarizing the state of the art, we can see that there are recurring claims that prosody conveys semantic and/or pragmatic information in English imperatives, but very little consensus about the details. With respect to whether there are prosodic morphemes in imperatives, Portner's and Rudin's analyses imply that, indeed, something akin to prosodic morphemes may be at play. Portner assigns different lexical semantics to 'falling' vs. 'rising' intonation, and Rudin assigns different discourse effects to H* L-L% vs. L* H-H%.[6]

Our examination of the prior literature suggests that there is a need for a more in-depth phonetic study on the prosody/meaning connection in English imperatives. Portner (2018), for example, as noted above, proposes a hypothesis about the semantic import of the intonational contours in falling and rising imperatives. Adopting an intuition put forth by Bolinger & Bolinger (1989), he suggests that 'steadily falling intonation is associated with commands and a rise with permission and invitation' (2018: 316). In earlier circulated versions of this paper, Portner further speculated that the rising intonation is based on an H phrase accent or a complex pitch accent, while the falling intonation involves an L accent. However, he admitted that the actual intonational analysis was still wide open. Rudin (2018) gives no phonetic measurements to support his claims about prosody, merely annotating 'rising imperatives' with a question mark. He explicitly states that '[t]he issue of how exactly a tune is translated into the continuous $f0$ across an utterance is taken to be a concern for the phonology-phonetics interface, and I provide no treatment of it here' (2018: 5). And although Chatzikonstantinou (2013) and Oikonomou (2016; 2022) did do phonetic measurements, their experiments were small pilot studies and no full results or statistics are reported.

We contribute to this body of work on the prosody and meaning mapping in imperatives through a series of three experiments on the prosodic distinction between command and non-command imperatives in English. As a reminder, our goals are to characterize the prosodic characteristics used by speakers and listeners to differentiate the two uses of imperative with the broader goal of elucidating the manner in which prosody-semantic interactions should be characterized.

Before we turn to our experiments, we have an important terminological note. For ease of exposition, we will be using the terms 'strong' and 'weak' for the two uses of imperatives we are investigating. It should be noted, however, that our 'weak' imperatives do not consist primarily of the weakest uses — permission — and therefore our 'weak' contexts are *not* primarily ones that would invite an analysis using a possibility modal (for those who prefer a modal analysis). Most of our 'weak' imperative contexts involve advice, and two of them are most naturally construed as permission. What was crucial for us in constructing the contexts was that the 'weak' ones are

---

[6] Oikonomou (2016) does not assign semantics to tonal melodies, claiming that the prosodic difference is ordinary F(ocus)-marking attracting the nuclear stress.

cases where the speaker believes the proposed action will be a priority for the *addressee*, while the 'strong' contexts are ones where the proposed action is a priority for the *speaker* (cf. Portner's 2018 characterization, discussed above).

4 Experiment 1: Perception from "idealized" production

We formed the hypothesis that listeners will assign imperatives produced with final L* L- L% (a low accent with a low fall at the phrasal boundary) contours to command discourse contexts, and imperatives produced with final H* L- L% (a high accent and low fall at the phrasal boundary) contours to advice discourse contexts. Our hypothesized contours were adapted from those proposed by Portner (2018), and matched our intuitions about the most likely contours for younger speakers in our geographical area. We then prepared materials for a perception task with these intonation contours.

4.1 Materials

Nineteen imperative sentences that were maximally voiced (containing few voiceless sounds to facilitate f0 estimation) were drafted. All of the imperatives invited positive actions on the part of the addressee; none contained negation or prohibitions. Using these imperatives, 38 contexts were designed in which we manipulated whether the speaker is in a position of equal vs. superior authority compared to the addressee, and whether the content of the imperative is the speaker's priority or the addressee's.

For each imperative, we created one command context and one advice context. In order to make the contextual contrast between the command interpretations and the advice interpretations maximally clear, we manipulated not only the discourse contexts but also the interlocutor roles. Thus, in the command contexts, the priority was the speaker's and moreover there was unequal authority between speaker and addressee such that the speaker was more powerful (e.g. the speaker is a head chef who wants a task completed, so they tell a subordinate to do it), while in the advice contexts, the priority was the addressee's and moreover there was equal authority (no power difference) (e.g., the speaker's friend wants help preparing dinner, and doesn't know where to start).

To ensure that our contexts were indeed interpreted as 'strong' or 'weak' by participants and to identify any ambiguous contexts for removal, we conducted a pre-experiment. This consisted of a survey containing all of our contexts, followed by a multiple-choice question presenting two rephrasings of the imperatives as containing either strong or weak necessity modals (e.g., "Jessica must boil the water" vs. "Jessica should boil the water") and the option to write in a third response if neither seemed a suitable paraphrase.

Fifteen volunteers completed the survey. On the basis of their responses, we discarded three imperatives, and rephrased certain contexts (e.g., making authority relations more clear). This left 16 imperatives, which are presented in **Table 1**.

Boil the nine litres of water for the ravioli.	Have a banana.
Buy Grandma a bouquet of roses.	Hide the money in the drawer.
Climb the mountain in the summer.	Move the wardrobe.
Divide the total by eleven.	Order dinner for Emily.
Do your homework before dinner.	Play an original arrangement.
Draw the girl using the brown crayon.	Review the order before finalizing.
Go home and weed the garden.	Roll the dough for the perogies.
Grill the buns and the burgers.	Throw the ball to Mario.

Table 1: Full list of target sentences used in experiments.

Cartoon images were designed to accompany the contexts, to make the intended meaning robust. **Figure 1** presents the example images that accompanied "Have a banana", and the strong and weak contexts for this string are given below in (9)–(10). All of the contexts are available at www.osf.io/5pm3h.

Figure 1: Accompanying images for the strong (top panel) and weak (bottom panel) contexts for the imperative "Have a banana."

(9)   *Strong context for "Have a banana":*
       Your daughter is suffering from some health problems that are mostly mysterious, but linked to potassium. According to the doctor she has to eat a banana every six hours or risk serious health consequences; cookies and sweets are off-limits until they've done more tests. Your daughter is whining about having to eat so much fruit. You are getting a little impatient with her and you say: Have a banana.

(10)   *Weak context for "Have a banana":*
       You and your friend are getting ready to go on a bike ride. Your friend is wondering whether she should have a quick snack before you head out. You just happened to have purchased some bananas, so you say: Have a banana.

One of the authors produced the imperative sentences in **Table 1** using the final L*L-L% or H*L-L% contours hypothesized as idealized examples of strong and weak imperatives. We recorded several productions of each target sentence in each prosodic condition using an AKG C250 head-mounted microphone and a SoundDevices 2.0 pre-amp. Recordings were made at 44.1kHz with a 16 bit-depth and saved directly as .wav files. The authors then selected the best version of each imperative to be used for the experiment. The stimuli are available for listening here: www.osf.io/5pm3h.

4.2 Procedure

Participants were individually tested on custom-built PC desktop computers in sound-attenuated cubicles. The experiment was administered using ePrime and participants wore AKG headphones. The stimuli were simultaneously presented to participants as written text and a contextualizing cartoon, followed by the audio recordings of the target imperative in both prosodic conditions (L*L-L% and H*L-L%). Listeners were asked to determine which production was more appropriate or natural for the context. Participants selected their preferred contour, responding to the prompt: "Judge which response is better given the situation." Responses were recorded using a serial response box.

Each context was presented twice through the course of the experiment, counterbalancing the presentation order of the strong and weak imperative response options. The order of the contexts was fully randomized for each participant.

Note that we also conducted two additional experiments not reported here: one which used only auditory prompts for the contexts (recorded by another trained linguist) instead of the written contexts, and another which used the auditory prompts for the contexts accompanied by the illustrations. There were no statistically significant differences amongst the three conditions

in a mixed effects logistic regression model, and thus for simplicity's sake, we simply report on the most basic of the designs.

4.3 Participants

Twenty-three self-reported native speakers of English with no current speech or language disorders or hearing impairments completed the task. Participants were recruited through a participant pool of students at the University of British Columbia, and were compensated with partial course credit.

4.4 Results

Responses made in under 250 ms were removed, which eliminated less than 1% of the data. Responses made more than two standard deviations away from the mean (M = 2091 ms, SD = 858) were also removed, which left over 95% of the original data set. The remaining data were fit to a mixed effects logistic regression model. The dependent measure was the proportion of trials that matched the strong and weak labeling of our target stimuli. Condition (strong, weak) was a fixed effect, and was dummy coded with strong as the reference level. Subject and Item (the 16 Response Text options) were entered as random slopes with Condition as a random intercept for each.[7]

The model reported a significant intercept [B = 1.48, SE = 0.31, z = 4.78, p < 0.001], which indicates that when presented with the strong contexts, listeners made judgments that supported our hypothesis. That is, listeners categorized the L*L-L% items as strong imperatives. The effect of Condition was not significant [B = –0.39, SE = 0.27, z = –1.46, p = 0.15], indicating that listeners were not significantly more accurate on strong trials compared to weak trials. Listeners associated the H*L-L% contour with weak imperatives and L*L-L% with strong imperatives. The listener means for strong and weak imperatives were both significantly above chance [Strong: M = 76% matched prediction, SE = 0.03, t(22) = 7.7, p < 0.001; Weak: M = 72% matched prediction, SE = 0.03, t(22) 6.39, p < 0.001]. These patterns are visualized in **Figure 2**. Overall, these results provide support for the hypothesis that the L*L-L% contour indexes strong imperatives, while the H*L-L% contour indexes weak imperatives, though listener responses are far from 100% matching the predictions.

We also conducted an analysis where we treated the data as a signal detection task (MacMillan & Creelman 2005). This method of analysis allows for the consideration of types of wrong responses (e.g., overapplication of the label "strong" for weak items or incorrectly

---

[7] The following code was used: glmer(ProportionMatch ~ Condition + (1 + Condition|Subject) + (1 + Condition|Item), family = "binomial").

calling a strong item weak), which allows us to examine biases and response strategies. For this analysis we arbitrarily labelled correct responses to strong items as hits in our calculation of d', a measure of sensitivity. Due to the direction of calculation, this means that negative c values, which is a measure of response bias, indicate a bias to respond "strong". Overall d' values were high, corroborating the accuracy analysis, demonstrating that listeners were good at the task (M = 2.7, SD = 1.4). Overall, listeners' bias (c) scores did not differ from 0 (M = –0.08, SD = 0.24; t(22) = –1.7, p = 0.10), suggesting that, overall, listeners were unbiased in their responses, indicating that they did not have a default assumption that imperatives are strong or weak.
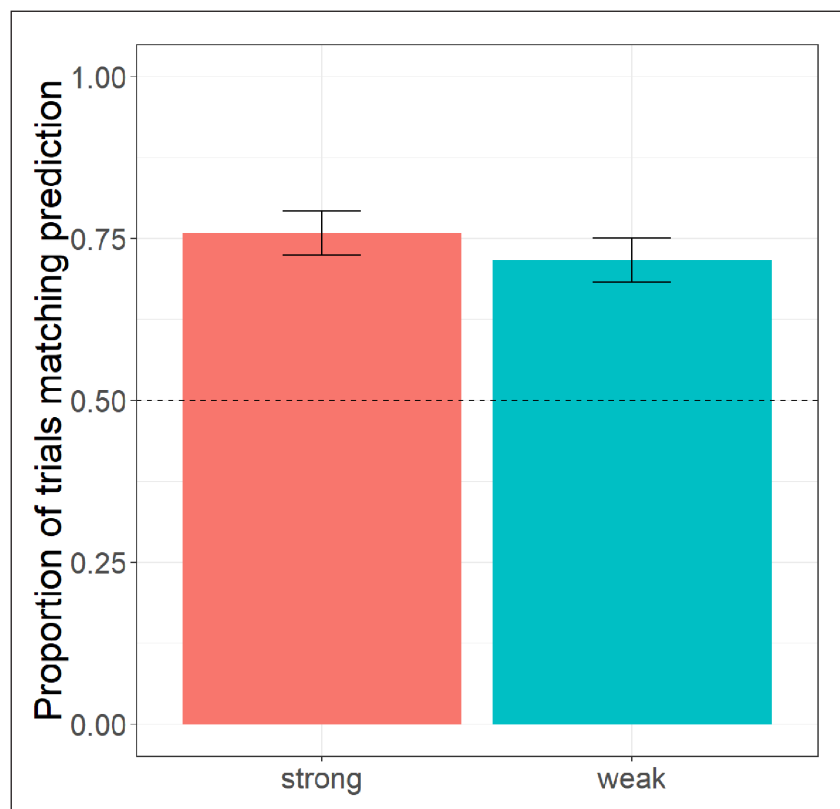
Figure 2: Proportion of trials matching the prediction for Strong and Weak contexts for Experiment 1. The error bars represent standard error.

Listener data are shown in **Figure 3**, where the left panel shows d' scores and the right panel shows c scores, ordered by listeners' values. These visualizations demonstrate that listeners' sensitivity to the strong and weak distinction varied considerably (min d' = 0.38, max d' = 5.04). Moreover, while overall listener bias was not significantly different from 0, bias scores ranged widely, with more participants (n = 13) showing a bias to respond strong than weak

(n = 9). Note that one participant fell right at the zero line, indicating an unbiased response strategy.
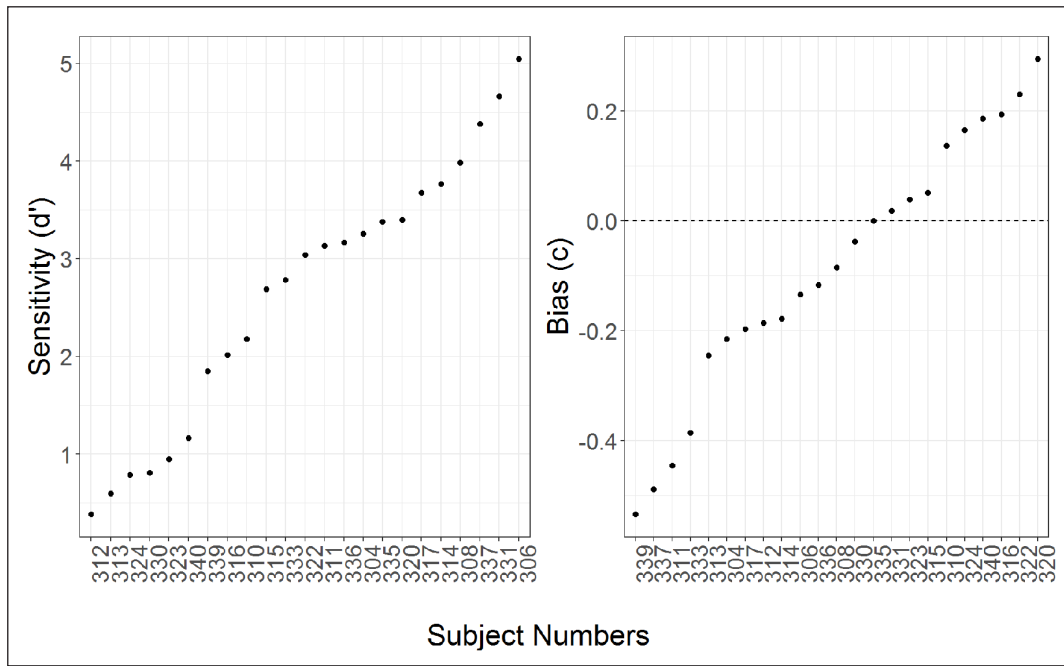
Figure 3: By-listener sensitivity (left panel) and bias (right panel) to the Strong and Weak contrast in Experiment 1. The data are ordered by listener sensitivity and bias, respectively. In the bias panel, the dashed line is at 0, demarcating unbiased performance. Positive bias values indicate a bias to respond weak, negative values indicate a bias to respond strong.

4.5 Discussion

The goal of this first experiment was to confirm that listeners associated the idealized intonation contours produced by a trained linguist with strong and weak imperatives. Listeners indeed do make the associations in the expected direction approximately 75% of the time. While this is far from unanimous, it does suggest that the proposed intonation contours convey meanings, potentially suggesting statistical gradience between the prosodic structure and semantic meaning, such that while a meaning may be associated with a category a large percentage of the time, that association is not reliable in 100% of cases (Ladd 2014). Specifically, Experiment 1 provides evidence that listeners generally associate the idealized intonation contour L*L-L% with command imperatives and the idealized intonation contour H*L-L% with advice imperatives. However, note that we will complicate this suggested interpretation below.

Given that the stimuli used in the first experiment were idealized and produced by linguists, in the next experiment we document the prosodic cueing speakers naturally produce in response

to the contexts written to convey strong and weak imperatives. To this end, for our second experiment we elicited productions from a new group of participants in response to the contexts.

5 Experiment 2: Production

For the production experiment, we hypothesized that participants' recorded imperatives would be similar to the "idealized" productions used for Experiment 1; that is, that participants would produce strong imperatives with a L* nuclear accent, and weak imperatives with H*, and that both conditions would have a L- phrasal tone and a L% boundary tone.

5.1 Materials

The materials for the second experiment were the 16 target imperatives and the 32 contexts (16 strong, 16 weak) described above for Experiment 1.

5.2 Procedure

Participants were asked to read a context, and then were presented with the imperative without punctuation as a text response. Participants were asked to record the response in a way that feels "natural or normal given the situation," where the situation was the context that had been provided. Each context was presented twice, such that four versions of each phrase were collected from each participant (2 strong, 2 weak for each phrase).

Recordings were made through Audacity directly to a Windows PC at a 44.1kHz sampling rate at 16 bit-rate. Participants wore a head-mounted AKG C250 microphone connected to SoundDevices 2.0 pre-amp.

5.3 Participants

Nineteen self-identified native speakers of English completed this task. Participants were recruited through a participant pool of students at the University of British Columbia, and were compensated with partial course credit.

5.4 Analysis
5.4.1 ToBI Analysis

Four trained linguists coded each utterance in terms of (1) a high or low accent on the stressed word; (2) whether that high or low accent was followed by a rise or a fall, and (3) whether the final stressed word was the last word in the utterance. Example sound files were provided to the coders. The transcriptions were collected via a Praat script, with decisions (1) and (2) combined into single code (e.g., H tone on final stressed word, fall after) and coders were instructed to keep

the pitch tracker off to place the focus on the auditory impression of the signal. Coding was done on a speaker-by-speaker basis, and the coders were instructed to listen to all of the sound files from one individual before beginning that speaker's coding to familiarize themselves with that speaker's pitch range.

5.4.2 Acoustic Analysis

The utterances were force-aligned with the Montreal Forced Aligner (McAuliffe et al. 2017) so that transcriptions of words and phones (segments) were time-aligned with the recorded .wav file. A research assistant manually checked and adjusted silent pause positions and placement, adjusting the estimated word boundaries given by the forced aligner to improve alignment accuracy, as necessary.

In annotating for our targeted set of ToBI markers, we noticed that the precise intonation contours produced were extremely variable, both within and across speakers. Furthermore, even outside of those phrase-final elements we were unable to identify a concrete alternative tone pattern that grouped one imperative reading to the exclusion of the other. Given this, we reserved an examination of the intonation contours to take place after a perceptual analysis by a group of listeners (participants in Experiment 3), to confirm that listeners are able to reliably identify the productions as strong and weak (see Section 6). At the same time, auditory impressions suggested that instead of particular intonation contours, the strong or weak imperative distinction was reflected in two macro settings: overall speech rate and overall pitch setting. The prosodic settings of speech rate and overall pitch setting are analyzed here.

Speech rate was calculated as the number of syllables per second in each utterance, where those syllables are estimated from the dictionary pronunciation of each word. This speech rate calculation includes (within-utterance) pauses in the calculation, as they form part of the utterance duration. To calculate pitch setting, f0 was estimated across the utterance in millisecond timestamps using the Straight algorithm (Kawahara et al. 1999) in VoiceSauce (Shue et al. 2011). On a by-speaker basis, f0 values were normalized to semitones using each speaker's 5% quantile as the baseline value.

5.5 Results

A subset of five speakers were analyzed upon two raters' completion of the annotation. Cohen's Kappa was calculated using the irr package (v. 0.84.1, Gamer et al. 2015) with these two raters for the tone on the stressed word, plus fall or rise categorization and the decision as to whether or not the stressed word was utterance final. Cohen's Kappa was used to quantify agreement. Moderate agreement was found both in whether there was a high or low accent on the stressed word and whether that high or low accent was followed by a rise or fall [$z = 9.17$, Kappa =

0.395, p < 0.001], as well in the coding regarding whether the final stressed word was the final word in the utterance [z = 5, Kappa = 0.225, p < 0.001]. While these agreement levels are in the moderate range, given that they were on a subset of the speakers and a comparison of only two of the coders, it was determined the agreement was not high enough to have confidence in the auditory coding.

5.5.1 Speech Rate

Speech rate was used as the dependent variable in a linear mixed effects model with Condition (dummy coded, with strong imperatives as the reference level) as a fixed effect and speaker and item as random effects. Condition was included as a random slope for both speaker and condition (model syntax: SpeechRate ~ Condition + (1+Condition|Speaker) + (1|Item)). There was a significant intercept [B = 4.85, SE = 0.25, t = 19.31] and an effect of Condition:Weak (B = 0.71, SE = 0.16, z = 4.3], indicating that imperatives that were elicited in a weak context were produced at a faster speech rate than imperatives produced in response to a strong context. These results are visualized in **Figure 4**, which presents individual participants' mean speech rate for each utterance condition. Each participant is represented by a point in each column that is connected by a single line.
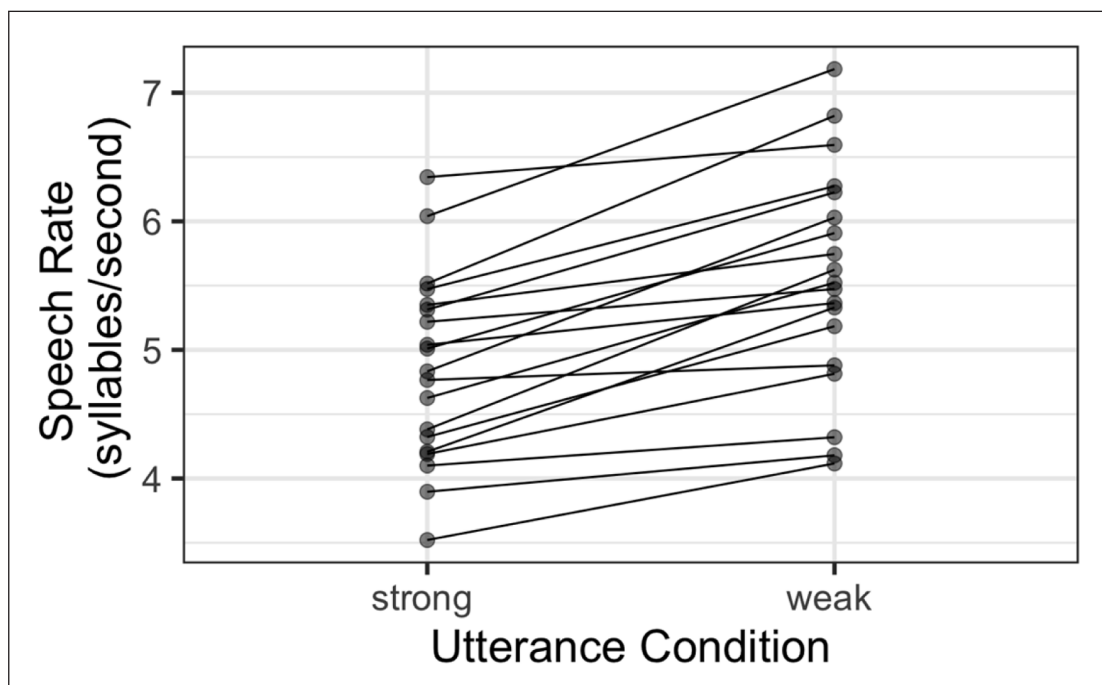
Figure 4: Speech rate (syllables/second) by the elicited imperative type. Each dot and connecting line represent an individual participant. Error bars are omitted for visual clarity.

5.5.2 Global pitch setting

Semitones were used as the dependent variable in a linear mixed effects model with Condition (dummy coded, with strong imperatives as the reference level) as a fixed effect and speaker and item as random effects. Condition was included as a random slope for both speaker and condition (model syntax: Semitones ~ Condition + (1+Condition|Speaker) + (1|Item)). There was a significant intercept [B = 8.05, SE = 0.59, t = 13.7] and an effect of Condition:Weak (B = 1.3, SE = 0.22, z = 5.973]. This indicates that weak imperatives had globally higher f0 settings than imperatives elicited in a strong pragmatic context; these results can be seen in **Figure 5**. Again, each participant is represented by a point in each column that is connected by a single line.
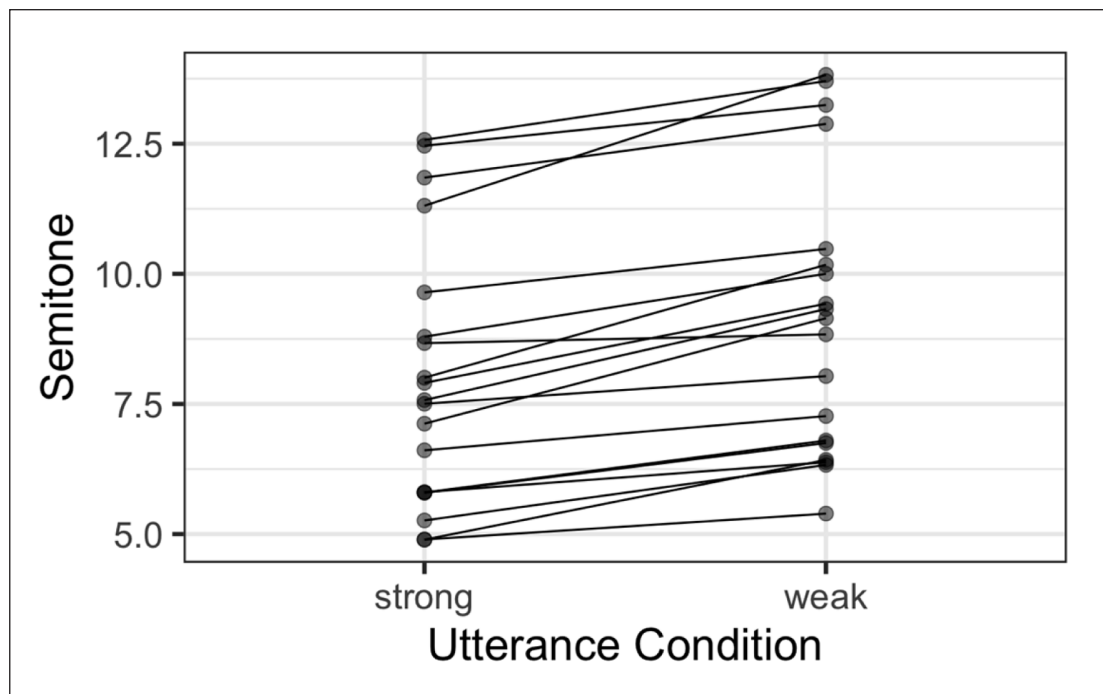
Figure 5: Semitones (normalized f0 estimates) by the elicited imperative type. Each dot and connecting line represents an individual participant. Error bars are omitted for visual clarity.

5.6 Discussion

ToBI transcription amongst the annotators did not provide an inter-rater reliability score that allowed for any sort of confident interpretation of the data. Our inability to reliably transcribe these utterances could be due to a lack of training, which is a noted challenge in prosodic transcription (Cole & Shattuck-Hufnagel 2016), or because of the actual physical gradience (Ladd 2014) or distributional realizations (Cangemi & Grice 2016) of the productions. We return to these points in the General Discussion (Section 8).

While the cross-speaker variability we observed prevented a clear prosodic description of these imperatives, two macro prosodic settings were both auditorily salient and acoustically robust. Speakers produced imperatives elicited in a weak context with a higher speech rate and a higher f0 setting than imperatives elicited in a strong context.

To see if listeners also make reliable associations of "low and slow" with strong imperatives, in the next experiment we presented an independent group of listeners with the elicited imperatives, asking them to categorize them as strong or weak.

6 Experiment 3: Perception from production

Auditory assessment by the authors indicates that the utterances gathered from participants in Experiment 2 showed substantial variation in how the imperatives were produced. To assess whether these productions are reliably interpreted as strong and weak imperatives, we presented them to listeners for categorization. Our hypothesis was that listeners can accurately interpret imperatives as 'strong' (command) or 'weak' (advice) based on hearing these natural productions, and that there would be a high degree of agreement between the original recording condition and listeners' judgments.

6.1 Materials

The imperative productions gathered from the 19 individuals described above in Experiment 2 were used in this task.

6.2 Procedure

Participants were individually tested on custom-built PC desktop computers in sound-attenuated cubicles. The experiment was administered using ePrime and participants wore AKG headphones. Listeners were presented with the imperatives over headphones and asked to categorize each imperative as a hard or soft command. Hard commands were defined for participants as "demanding, ordering, or requiring that the listener *must* follow the command", and soft commands were defined as "recommending, advising, or suggesting that the listener *might* follow the command."

Listeners were presented with three randomly selected speakers, presented as three separate blocks with self-timed breaks in between. Within each speaker block, each utterance was presented twice. The order of the utterances was randomized for each listener. Each speaker was presented to 14–18 listeners.

6.3 Participants

A total of 101 self-reported native speakers of English with no current speech or language disorders or hearing impairments or disabilities completed this task. Participants were recruited through a participant pool of students at the University of British Columbia, and compensated with partial course credit.

6.4 Perception results

Trials for which there were no responses were removed, eliminating less than 0.5% of the data. Responses made in under 250 ms were also removed, which accounted for less than 0.2% of the data. Then responses made more than two standard deviations away from the mean response time (M = 2044 ms, SD = 824) were trimmed, which eliminated another 4.5% of the data set. The data were fit to a mixed effects logistic regression model to predict the proportion of trials that matched the elicited production category. Category (strong, weak) was a fixed effect, and was dummy coded with strong as the reference level. Subjects, Speakers, and Items (the 16 imperatives) were entered as random slopes with Condition as a random intercept for each.[8]

The model returned a significant intercept [B = 1.34, SE = 0.20, z = 6.58, p < 0.001], which, given that strong was the reference level, indicates that listeners generally accurately identified strong imperatives as such. The effect of Condition was significant [B = –0.60, SE = 0.24, z = –2.51, p = 0.01]; listeners were less accurate at categorizing weak imperatives than strong imperatives. T-tests confirmed that listener performance was significantly above chance for both strong [M = 76%, t(100) = 35.06, p < 0.001] and weak [M = 66%, t(100) = 18.15, p < 0.001] imperatives. **Figure 6** illustrates these results visually.
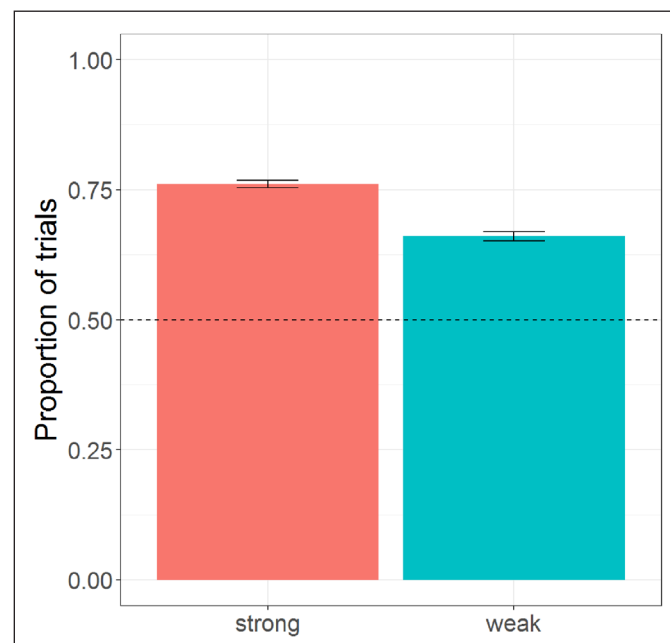
Figure 6: Proportion of trials matching the elicited productions for Strong and Weak contexts for Experiment 3. The error bars represent standard error.

---

[8] The following code was used: glmer(Accuracy ~ Category + (1 + Category|Subject) + (1 + Category|Speaker) + (1 + Category |Item), family = "binomial").

As for Experiment 1, the data were analyzed as a signal detection task (MacMillan & Creelman 2005) in two analyses: one which assessed the behaviour of the listeners, and another which quantified the imperative contrast robustness on a by-speaker level. As in the previous SDT analysis, strong items were coded as hits in our calculations. Listeners were generally quite good at the task (M = 2.4, SD = 0.46). Listeners showed a significant bias to respond strong (M = –0.15, SD = 0.17; t(100) = –8.91, p < 0.001, Cohen's d = 0.88). **Figure 7** presents listeners' individual performance in terms of d' (sensitivity) and c (bias). Listeners' sensitivity to speakers' strong and weak contrasts was generally high, but the majority of listeners showed a strong response bias, with one participant rather strongly biased towards a weak response.
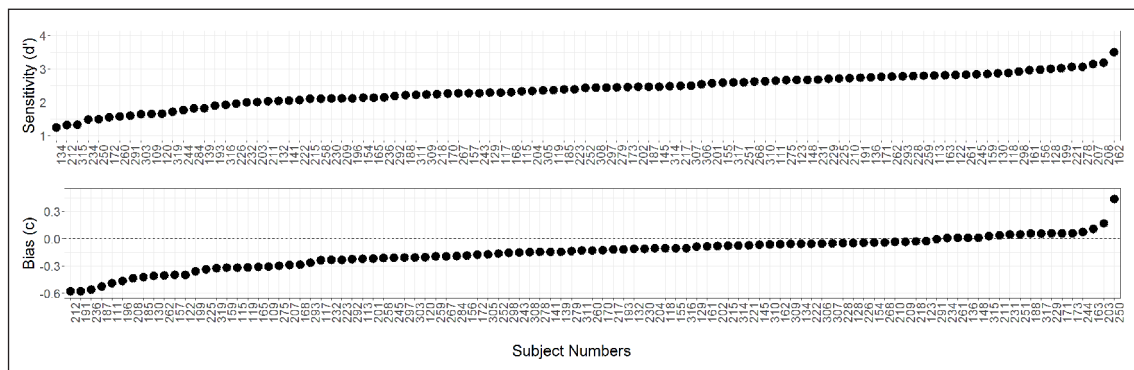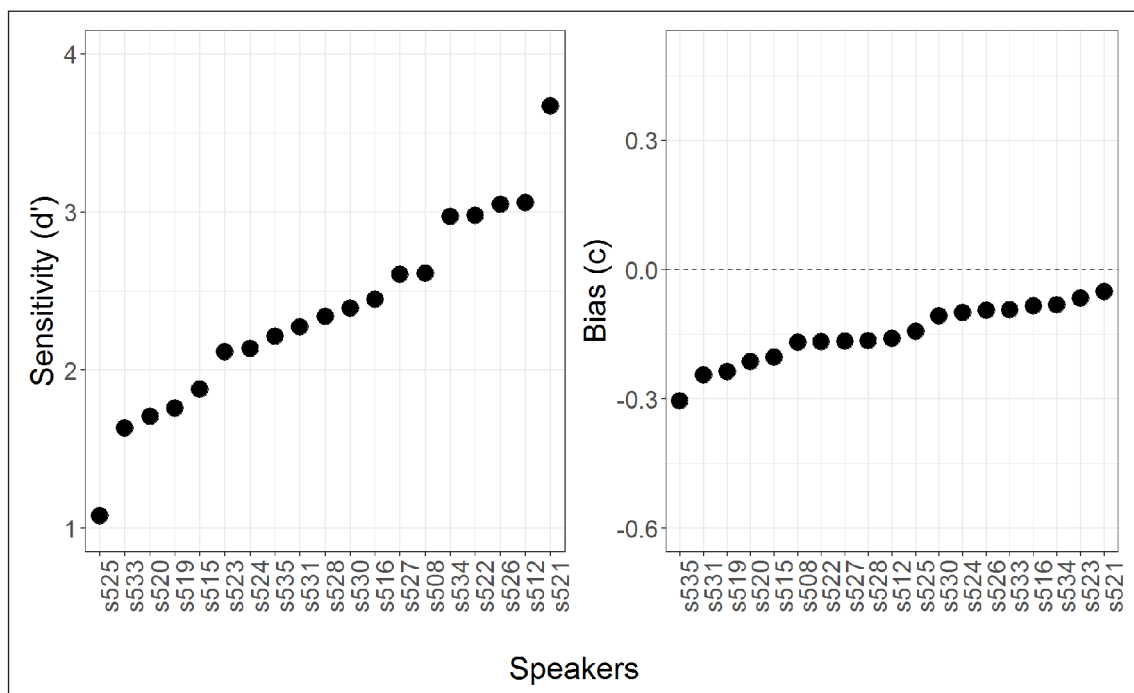

Figure 7: By-listener sensitivity (top panel) and bias (bottom panel) to the participant-produced Strong and Weak contrast in Experiment 3. The data are ordered by listener sensitivity and bias, respectively. In the bias panel, the dashed line is at 0, demarcating unbiased performance. Positive bias values indicate a bias to respond weak, negative values indicate a bias to respond strong.

Given that listeners were randomly assigned to three speakers, a listener's mean sensitivity is going to be highly affected by the robustness of the strong and weak contrast for the speakers to which they were assigned. To understand the speaker differences, d' and c were calculated on a by-speaker basis. On average, sensitivity for speakers was relatively high (M = 2.36, SD = 0.62). Speakers' voices also elicited a strong response bias that was significantly different from 0 [M = –0.15, SD = 0.07; t(18) = –9.56, p < 0.001, Cohen's d = 2.19). **Figure 8** presents these by-speaker values, ordered by sensitivity and bias. Speakers vary considerably in the robustness of their production of the contrast, but all speakers elicited a "strong" response bias.

6.5 Interim summary of all three experiments

Experiment 1 showed that listeners can categorize imperatives as strong or weak based on idealized productions of L* L- L% and H* L- L% respectively, fairly well. Speakers' strong and weak imperative productions from Experiment 2 were highly variable in terms of contours, but

with consistently higher speech rate and overall pitch for weak than for strong. Experiment 3 demonstrated that listeners can categorize imperatives as strong or weak based on natural productions fairly well, and that these accuracy levels from the spontaneously elicited utterances align well with the idealized utterances from Experiment 1 (e.g., mean accuracy on the strong imperatives was 76% in both Experiment 1 and Experiment 3, and the mean accuracy on the weak imperatives across the experiments was 72% and 66%, respectively).

Figure 8: By-speaker analysis of sensitivity (left panel) and bias (right panel). The data are ordered by listeners' average sensitivity and bias to each speaker. In the bias panel, the dashed line is at 0, demarcating unbiased performance. Positive bias values would indicate a bias to respond weak, negative values indicate a bias to respond strong.

This alignment of response accuracy across the two perception experiments, along with the observations from the production experiment that global pitch setting and speech rate were robust differences in the spontaneous imperative productions, prompted us to analyze the idealized productions from Experiment 1. Like the spontaneous productions, the strong (M = 13.8 semitones, SD = 3.4) and weak (M = 14.6 semitones, SD = 3.2) idealized productions also differed significantly in global pitch setting (t(15) = –3.15, p = 0.007, Cohen's d = 0.96). Unlike the spontaneous productions, however, the idealized productions did not differ in their speech rates between the strong (M = 5.8 syllables/second, SD = 0.7) and weak (M = 5.9, SD = 0.7) imperatives (t(15) = –0.87, p = 0.4).

Together, this suggests that listeners may be using a speaker's overall f0 setting as a prosodic dimension by which to determine the status of the imperative as semantically/pragmatically strong or weak.

7 Perception results as a guide

The results of the previous analysis indicate that, overall, listeners are able to identify speakers' imperative productions as reliably strong or weak in a way that aligns with the context in which the imperatives were produced. This is despite the fact that the intonation patterns of participants' productions were highly variable. The perception results, however, can be a guide to identifying which imperative productions indicated their strength most clearly. In this section we focus on the 7% of the utterances in the perception experiment which received unanimous (i.e., 100%) agreement from the listeners. We conservatively focus on these items, though agreement was high on a large number of items; 28% of the utterances were categorized in the intended direction by listeners 90% of the time.

7.1 Methods

Eighty-four utterances, constituting 7% of the items from Experiment 2, received unanimous agreement from listeners. Three of the authors listened to these items together and auditorily coded the items as follows. We coded for contours using auditory perception-based annotation only (i.e., we did not look at the Praat pitch traces). Nuclear accent annotation fell into three groups:

(11) 'high': nuclear accent was higher than syllables preceding and following (a "bump" followed by a fall)
'level then fall': nuclear accent was equal to syllable preceding and higher than syllable following
'low': nuclear accent was lower than syllables preceding and following

We also coded for final rise/final fall; the overwhelming majority of data (82/84) were coded as ending in a fall rather than a rise.

7.2 Results

The coding provided frequency counts that were analyzed using Pearson's Chi-squared test for the auditory coding of nuclear accent, subjective speech rate, and final rise.

7.2.1 Nuclear Accent

The nuclear accent was auditorily coded as high, level-then-fall, low, or N/A. The percentage of times that each of these labels was applied to the elicited strong and weak items is shown in

Table 2, with the raw frequency counts in parentheses. The N/A category was ignored and the frequency counts for the remaining categories were subjected to a Pearson's Chi-squared test, which established a difference across cells ($X(2) = 41.64$, $p < 0.001$). Given the small number of observations in the Low category, we focus our attention on the High and Level-then-Fall comparisons. Imperatives elicited in response to strong contexts and unanimously categorized as sounding "strong" by listeners were much more likely to have the Level-then-Fall nuclear accent than the High nuclear accent. Comparatively, the imperatives unanimously perceived as "weak" by listeners and elicited in the weak context were much more likely to have High nuclear accents than the Level-then-Fall designation.

	High	**Level then Fall**	**Low**	**N/A**
strong	10% (6)	72% (43)	7% (4)	12% (7)
weak	79% (19)	4% (1)	12% (3)	4% (1)

Table 2: Percentage (counts) of elicited strong and weak items receiving each coding category for Nuclear Accent

7.2.2 Subjective Speech Rate

The subjective speech rate of the utterances was coded as fast, normal, or slow. The percentages and raw frequency counts for this coding are shown in **Table 3**. The frequency counts were submitted to a Pearson's Chi-squared test, which found a difference across cells ($X(2) = 37.58$, $p < 0.001$). Imperatives elicited in weak contexts were more likely to be perceived as fast compared to the strong imperatives, which were more likely to be categorized as normal or slow. Notably, no weak imperatives were perceived as slow.

	Fast	**Normal**	**Slow**
strong	12% (7)	57% (34)	32% (19)
weak	79% (19)	21% (5)	0% (0)

Table 3: Percentage (counts) of elicited strong and weak items and their auditory coding with respect to perceived speech rate.

7.2.3 Final Rise

The presence or absence of a final rise was coded as present vs. absent. The percentages and counts of the coding by utterance type are reported in **Table 4**. The frequency counts were used in a Chi-square test, which found no difference across cells ($X(1) = 2.2$, $p = 0.14$).

	No	Yes
strong	100% (60)	0% (0)
weak	92% (22)	8% (2)

Table 4: Percentage (counts) of elicited strong and weak items and their auditory coding with respect to the presence of a final rise.

7.3 Discussion

The 'clearest' utterances — those which participants in Experiment 3 unanimously categorized according to the strong or weak context they had been uttered in — were coded differently in terms of the perceived nuclear accent and subjective speech rate when they were imperatives elicited in strong or weak contexts. Among this set of clearest utterances, strong imperatives were overall produced with a slower speech rate than weak imperatives, and the types of imperative also differed in the perceived height of their Nuclear Pitch Accent: while weak imperatives overwhelmingly had a H* accent that was both preceded and followed by syllables with lower pitch, the great majority of strong imperatives had a 'level then fall' contour, in which the Nuclear Pitch Accent was approximately level to the preceding syllable, but higher than the following one.

These results may appear to suggest that the clearest utterances had consistently different tonal contours after all, in spite of what we concluded based on Experiment 2. However, we do not believe that these strong and weak imperatives, which were the clearest in terms of recognition by participants, involve different tonal melodies. Notice that 'level-then-fall' is still a falling tune, which therefore would still be annotated as H* L- L%. The H* in a level-then-fall contour is merely produced without an extra high bump. Instead, it participates in the global pitch declination of the utterance. This may relate to Chatzikonstantinou's (2013) finding in his pilot study that strong imperatives had a "flat intonation following H H L- (L%) contour." Altogether, this would suggest that the final prosodic melody in both strong and weak imperatives is a falling contour. The difference lies in the phonetic implementation of the nuclear pitch accent, which is higher for weak imperatives. That is, weak and strong imperatives share the phonological structure for the nuclear pitch accent, that may vary in likelihood of downstep (likely for strong, not probable but plausible for weak), but they differ in terms of the phonological structure with respect to the macro prosodic characteristics that robustly distinguish these opposing semantic meanings.

8 General discussion

We have made some progress towards answering our original four research questions, repeated in (12).

(12)   i.   Do speakers produce 'command' and 'advice' imperatives with different prosody?
       ii.   If yes, what are the prosodic characteristics of the different types of imperatives?

iii. Do listeners use these production cues in their perception in order to disambiguate between different types of imperative?

iv. How should prosody-semantics interactions in imperatives be analyzed?

Experiment 2 provided evidence that English speakers indeed do produce 'command' and 'advice' imperatives differently, but these meanings were not simply delivered via tunes. The prosodic signature of these imperative meanings was robustly manifested as macro phonetic settings: speakers produce strong imperatives with a slower speech rate and a lower overall pitch than weak imperatives. In terms of intonational contours, we found substantial variability both within and between speakers, which impeded quantitative analysis of acoustic characteristics.

The evidence that listeners make use of production cues to disambiguate different types of imperatives comes from both Experiments 1 and 3. Listeners were fairly good at correctly categorizing imperatives as strong or weak, when hearing either idealized productions (Experiment 1) or natural ones (Experiment 3). This accuracy was in spite of the substantial within- and across-speaker variability observed in the participants' productions in Experiment 2, which were the stimuli in Experiment 3.

The idealized stimuli in Experiment 1, which were designed to test the hypothesis that the strong imperative meaning is manifested as L*L-L% and the weak imperative as H*L-L%, elicited accurate responses from listeners (76% accurate for strong imperatives, 72% for weak). Subsequent analysis revealed, however, that the strong and weak imperatives were also differentiated by overall f0 setting: strong imperatives were associated with a lower f0 than weak imperatives. Listeners' productions of strong and weak imperatives did not consistently exhibit the predicted prosodic contours, but they did robustly differ in terms of f0 setting and speech rate. Strong imperatives were low and slow, while weak imperatives had higher f0 and faster speech rates. There was variability across listeners in Experiment 3, and, notably, listeners were more accurate with the strong imperatives (76% accurate) than weak (66% accurate). We refer to the f0 and speech rate differences as distinctions in macro prosodic settings.

Despite the variability in production, many items in Experiment 3 were unanimously correctly categorized by listeners; these items were analyzed further. In these clearest utterances, which were uniformly and accurately categorized by all listeners, speakers not only consistently produced strong imperatives 'slower and lower' than weak ones, they had fairly consistent patterns for the height of the Nuclear Pitch Accent, based on our auditory coding. This consistency was particularly evident for the strong imperatives, for which the level-then-fall nuclear pitch accent pattern was very consistent (**Table 2**). The pattern of the high nuclear pitch accent for the weak imperatives was robust, but not as consistent as that for the strong imperatives. Ultimately, however, we propose that these contours should be analyzed identically: both as falls. The difference is in the phonetic height of the H* accent, which is higher on weak imperatives.

Our postulation that the strong and weak imperative distinction does not correspond to a consistent difference in phrase-final intonational melodies leads us to re-examine the conclusions we drew from Experiment 1, where we had hypothesized that it was the idealized pitch contours L* L- L% vs. H* L- L% that listeners were using as cues. As noted above, upon analysis we discovered that the idealized productions for Experiment 1 actually had higher overall pitch for weak imperatives than for strong ones. This may suggest that listeners in Experiment 1 were categorizing imperatives as strong or weak primarily using overall f0 setting, rather than the idealized prosodic contours we were aiming for. This is an interpretation that can be tested in future work.

What light do these findings shed on the final research question: how best to analyze the prosody-meaning connection in imperatives? Recall from Section 2 that several authors have proposed, in essence, prosodic morphemes consisting of particular tonal contours that convey particular meanings. Our results cast doubt on the validity of the assumption that final tonal contours are what categorically distinguish imperative meanings, at least for command vs. advice imperatives in the variety of English we investigated. Nevertheless, we did find prosodic differences between command and advice imperatives that are perceived and (presumably) used by listeners to distinguish the two types of speech act.

There seem to be two routes to take with respect to the interpretation of our results.[9] On the one hand, we could propose prosodic morphemes that convey meaning but that do not involve intonational melodies such as falls or rises. Instead, they involve the macro settings of overall pitch height, speech rate, and the height of the nuclear pitch accent. Under such an approach, and taking Portner's (2018) semantic analysis as an example for concreteness, we could say that lower overall pitch, slower speech rate, and a level-then-fall nuclear pitch accent encode that the speaker is committed to the property denoted by the imperative being a priority, while higher overall pitch, faster speech rate, and a high nuclear pitch accent encode that the speaker is asking for confirmation that the addressee is committed to the action.

As with any phonetic feature, we expect substantial variability in how these phonological instructions of 'high and fast' vs. 'low and slow' and the associated pitch accents are actually produced by speakers. While cross-speaker variation is always to be expected, within this particular meaning space, individuals will also differ substantially in their socio-pragmatic approaches to these imperatives. That is, individuals will deliver command and advice imperatives

---

[9] An important aside is that the theoretical and empirical scholarship on the prosody-meaning connection can be broadly separated into the direct and indirect camps. The indirect approach posits phonological prosodic categories produced as particular phonetic characteristics that mediate the mapping from semantics to acoustics. This is the intonational phonology approach (e.g., Gussenhoven 1983; Ladd 1996/2008). The direct approach posits that acoustic characteristics are directly associated with semantic meaning (e.g., Xu & Xu 2005). Our data do not adjudicate between these approaches. We interpret our data following the intonational phonology or indirect approach (Ladd 1996), where a phonological category mediates the mapping between semantic meaning and phonetic realizations.

with different degrees of politeness, the acoustic realization of which may muffle the clarity and preciseness of the originally targeted prosodic morpheme as the semantic and socio-pragmatic meaning is delivered through the same acoustic channel.

As an alternative to that first route, we could argue that command and advice imperatives are phonologically identical, and therefore that there are not two different form-meaning mappings, and only one phonological form. This would mean that there is semantic and pragmatic ambiguity between the two types of speech act. Under such an approach, the different phonetic realizations of commands vs. advice are merely ways that speakers use phonetic cues to attempt disambiguation between the two meanings.

Under the first approach, there would be a grammatical link between the meaning of a strong imperative and 'low and slow' productions, and between the meaning of a weak imperative and 'high and fast' productions. Under the second approach, in contrast, the connection between the meaning of the two types of imperatives and their phonetic productions is fully realized through the socio-pragmatic differences in how individuals approach imperative as a speech act. That is, the social-pragmatic meaning of the speech act is derived from a particular speaker's phonetic production. A concrete example of how this second approach could function would be as follows: low and slow speech patterns are associated with individuals who have or claim authority in social interactions. Speakers can then reproduce these low and slow features to signal a strong interpretation, leveraging that social association.[10] Regardless of the approaches followed, the phonetic distinctions we observed could arise as instantiations of more general pragmatic phenomena such as the speaker's level of authoritativeness, assertiveness, politeness, formality, or the relative positions of speaker and listener on a social hierarchy. This is similar to Oikonomou's (2016) suggestion that certain prosodies applied to imperatives can signal uncertainty or degree of endorsement.

Considering the second analytical option, recall that we found that weak imperatives are produced with a higher f0 setting than strong imperatives. This was true both in the idealized stimuli produced by a trained linguist for Experiment 1 and in the elicited productions from speakers in Experiment 2. This carries a flavour of Ohala's Frequency Code, which leverages sexual dimorphic generalizations and binary gender-stereotypic associations to predict that higher-pitched signals cue uncertainty versus the certainty indicated in a lower-pitched register (Ohala 1983). Strong imperatives can be viewed as being associated with authority, confidence, and certainty, while weak imperatives are not. Something like the Frequency Code might therefore provide the indirect link to explain the f0 differences found in our experiments between command and advice imperatives. However, a successful account of the prosodic differences between strong and weak imperatives cannot solely rest with an explanation like Ohala's

---

[10] Thank you to a reviewer for this clear characterization.

Frequency Code, given that imperative type was also robustly associated with speech rate, and there is little evidence that speech rate aligns with the Frequency Code. While it is certainly a stereotype within English-speaking communities that women speak faster than men, there is no evidence supporting such a claim (e.g., Byrd 1992). Women are perceived as speaking faster than men, but this is possibly due to traversing acoustically larger vowel spaces with, on average, smaller physiology (Weirich & Simpson 2014).

As noted in the Discussion of Experiment 2, the challenge to reliably transcribe the utterances elicited from participants in Experiment 2 could be due to the physical (Ladd 2014) or distributional (Cangemi & Grice 2016) gradience of realizations. The challenge could also be due to speakers' individual differences in the tonal realizations of strong and weak imperatives, as has been noted in the production of pitch accents (e.g., Niebuhr et al. 2011). Politeness is also a dimension that can affect the prosodic realization of an utterance (Cangemi & Grice 2016; Jeong & Potts 2016), and some participants may have adopted a more polite style in the weak imperatives that altered the expected prosodic contour.

Given this, and given the robustness of the macro-setting effects we found in addition to the differences in nuclear pitch accent height observed for the clearest of utterances, we lean towards the former type of analysis, namely that there is a grammatical/phonological link connecting a piece of semantics to a piece of phonetics. As **Figures 4** and **5** above demonstrated, every participant in Experiment 2 produced command imperatives both lower and slower than they produced advice imperatives. In spite of there being substantial variability in other phonetic properties of the productions, there seems to be a very clear generalization that commands are not produced high and fast, and advice is not produced low and slow. Moreover, these phonetic differences between command and advice imperatives are robustly used by listeners to disambiguate meaning. As noted above, although only 7% of the items produced in Experiment 2 were correctly categorized by the listeners in Experiment 3 with 100% accuracy, 28% of the items were categorized in the intended direction by listeners with 90% accuracy.

Recall that the predominant nuclear pitch accent contour we coded for the 'clearest' strong productions was 'level-then-fall', which is a common contour for declaratives. The rise-before-fall we predominantly heard in weak imperatives is the unexpected case intonationally. This could suggest that the strong interpretation is more basic, and that extra information or help is needed to convey the weak interpretation. There is some evidence in support of this in our data, as listeners were more likely to show a bias to respond "strong" to any imperative.[11] Similarly,

---

[11] A reviewer suggests an alternative to consider, namely that the weak interpretation is more basic, but adults have a tendency to derive strengthened interpretations when they can (similar to the strengthening of 'or' to an exclusive interpretation from an inclusive one). This is an interesting idea to pursue in future research. We do not adopt it at this time because we are not committed to a semantic analysis of command vs. non-command imperatives that places them on a scale of semantic strength.

future research might uncover some clues about the default interpretation of imperatives from our pitch and speech rate findings. That is, future experiments could probe whether the baseline speech rate and overall pitch for speakers are closer to those we found for the strong imperatives, or for the weak.

Our inquiry into the phonetic space of command and advice imperatives confirms that Portner (2018) was correct about the semantic distinction being realized phonetically. Our data, however, suggest he was incorrect about the nature of the phonetic manifestation of the semantic categories. The phonetic distinction between command and advice imperatives, while reliably detectable by listeners, is not wholly determined by final contour differences, but is rather signaled multidimensionally by speech rate, global pitch settings, and nuclear pitch accent. This constellation of features is not surprising, as it has long been noted that prosodically cued meaning goes beyond pitch contours (e.g., Gobl & Chasaide 2003; Campbell & Mokhtari 2003; Steedman 2014), but it does serve as a reminder of the complex mapping between meaning and form.

Data availability

Materials we have permission to share are available here: www.osf.io/5pm3h.

Ethics and consent

This research was approved by the Behavioural Ethics Research Board at the University of British Columbia in accordance with the ethical standards documented in the 1964 Declaration of Helsinki and its later amendments.

Funding information

This research has been supported by a SSHRC Insight Grant to MB.

Acknowledgements

Thank you to Jobie Hui for her artistry in the drawings for the experiment. Thank you to Lydia Rhi, Ariana Zattera, Brianne Senior for their contributions to data collection and analysis. Thank you to Khushi Patil and Megan Housley for assistance in manuscript preparation. For feedback, thank you to Michael Rochemont, Hotze Rullmann, Hubert Truckenbrodt, members of the Speech in Context Lab, and audiences at LabPhon 15 at Cornell University, the University of British Columbia, and the Universität zu Köln.

Competing interests

The authors have no competing interests to declare.

References

Bartels, Christine. 2014. *The intonation of English statements and questions: A compositional interpretation.* New York: Routledge. DOI: https://doi.org/10.4324/9781315053332

Bhadra, Diti. 2020. The semantics of evidentials in questions. *Journal of Semantics* 37(3). 367–423. DOI: https://doi.org/10.1093/jos/ffaa003

Bolinger, Dwight & Bolinger, Dwight L. M. 1989. *Intonation and its uses: Melody in grammar and discourse.* Stanford University Press. DOI: https://doi.org/10.1515/9781503623125

Byrd, Dani. 1992. Preliminary results on speaker-dependent variation in the TIMIT database. *The Journal of the Acoustical Society of America* 92(1). 593–596. DOI: https://doi.org/10.1121/1.404271

Campbell, Nick & Mokhtari, Parham. 2003. Voice quality: the 4th prosodic dimension. *Proceedings of the 15th International Congress of the Phonetic Sciences* (ICPhS, 2003), 2417–2420.

Cangemi, Francesco & Grice, Martine. 2016. The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Laboratory Phonology* 7(1). 1–20. DOI: https://doi.org/10.5334/labphon.28

Carter, Sam. 2021. Force and choice. *Linguistics and Philosophy* 45. 875–910. DOI: https://doi.org/10.1007/s10988-021-09335-w

Charlow, Nathan. 2010. Restricting and embedding imperatives. In Aloni, M. & Bastiaanse, H. & de Jager, T. & Schulz, K. (eds.), *Logic, Language and Meaning. Lecture Notes in Computer Science* 6042. Berlin, Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-642-14287-1_23

Charlow, Nathan. 2011. *Practical language: Its meaning and use.* University of Michigan Doctoral dissertation.

Chatzikonstantinou, Tasos. 2013. *Prosody and illocutionary force in English imperatives.* Ms., University of Chicago.

Cole, Jennifer & Shattuck-Hufnagel, Stefanie. 2016. New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology* 7(1). DOI: https://doi.org/10.5334/labphon.29

Condoravdi, Cleo & Lauer, Sven. 2012. Imperatives: Meaning and illocutionary function. In Piñon, Christopher (ed.), *Empirical issues in syntax and semantics* 9, 1–21. http://www.cssp.cnrs.fr/eiss9/eiss9_condoravdi-and-lauer.pdf

Davis, Christopher. 2009. Decisions, dynamics and the Japanese particle yo. *Journal of Semantics* 26(4). 329–366. DOI: https://doi.org/10.1093/jos/ffp007

Fletcher, Janet & Loakes, Deborah. 2010. Interpreting rising intonation in Australian English. In *5th International Conference of Speech Prosody* 2010.

Gamer, Mattias & Lemon, Jim & Singh, Ian F. P. 2015. *Various coefficients of interrater reliability and agreement.* R package version 0.84. 2012.

Gobl, Christer. & Chasaide, Ailbhe N. 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech communication* 40(1–2). 189–212. DOI: https://doi.org/10.1016/S0167-6393(02)00082-1

Grosz, Patrick. 2009. German particles, modality, and the semantics of imperatives. In Lima, Suzi & Mullin, Kevin & Smith, Brian (eds.), *North East Linguistic Society* (NELS) 39, 323–336.

Gunlogson, Christine. 2001. *True to form: Rising and falling declaratives as questions in English.* New York, NY: Routledge. DOI: https://doi.org/10.4324/9780203502013

Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22. 101–36. DOI: https://doi.org/10.1075/bjl.22.06gun

Gussenhoven, Carlos. 1983. Focus, mode and the nucleus. *Journal of Linguistics* 19(2). 377–417. DOI: https://doi.org/10.1017/S0022226700007799

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

Han, Chung-hye. 1999. Deontic modality, lexical aspect and the semantics of imperatives. *Linguistics in the morning calm* 4. 475–95.

Han, Chung-hye. 2019. Imperatives. In Portner, P. et al. (eds.), *Semantics - Sentence and Information Structure*, 225–249. Berlin, Boston: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110589863-006

Heim, Johannes M. 2019. *Commitment and engagement: The role of intonation in deriving speech acts*. University of British Columbia Doctoral dissertation.

Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Jeong, Sunwoo & Condoravdi, Cleo. 2017, February. Imperatives with the calling contour. In *Proceedings of Berkeley Linguistic Society* 43.

Jeong, Sunwoo & Condoravdi, Cleo. 2018. Imperatives and intonation: The case of the down-stepped level terminal contour. In Bennett, Wm. G. et al. (eds.), *Proceedings of the 35th West Coast Conference on Formal Linguistics*, 214–223.

Jeong, Sunwoo & Potts, Christopher. 2016. Intonational sentence-type conventions for perlocutionary effects: An experimental investigation. In *Semantics and Linguistic Theory* 26, 1–22. DOI: https://doi.org/10.3765/salt.v26i0.3787

Kaufmann, Magdalena. 2012. *Interpreting imperatives*. Dordrecht: Springer. DOI: https://doi.org/10.1007/978-94-007-2269-9

Kaufmann, Magdalena & Kaufmann, Stefan. 2016. Modality and mood in formal semantics. In van der Auwera, Johan & Nuyts, Jan (ed.), *The Oxford handbooks of mood and modality*, 785–820. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199591435.013.24

Kawahara, Hideki & Katayose, Haruhiro & Cheveigné, Alain D. & Patterson, Roy D. 1999. Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. In *Sixth European Conference on Speech Communication and Technology*. DOI: https://doi.org/10.21437/Eurospeech.1999-613

Ladd, D. Robert. 1978. Stylized intonation. *Language* 54(3). 517–540. DOI: https://doi.org/10.1353/lan.1978.0056

Ladd, D. Robert. 1996/2008. *Intonational Phonology*. 1st and 2nd edition. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511808814

Ladd, D. Robert. 2014. *Simultaneous structure in phonology* (Vol. 28). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199670970.001.0001

Macmillan, Neil & Creelman, C. Douglas. 2005. *Detection theory: a user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malamud, Sophia A. & Stephenson, Tamina. 2015. Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics* 32(2). 275–311. DOI: https://doi.org/10.1093/jos/ffu002

McAuliffe, Michael & Socolof, Michaela & Mihuc, Sarah & Wagner, Michael & Sonderegger, Morgan. 2017, August. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Interspeech* 2017, 498–502. DOI: https://doi.org/10.21437/Interspeech.2017-1386

Niebuhr, Oliver & d'Imperio, Mariapaola & Fivela, Barbara G. & Cangemi, Francesco. 2011. Are there "shapers" and "aligners"?: Individual differences in signaling pitch accent category. In *The International Congress of the Phonetic Sciences,* 120–123.

Ohala, John. J. 1983. Cross-language use of pitch: An ethological view. *Phonetica* 40(1). 1–18. DOI: https://doi.org/10.1159/000261678

Oikonomou, Despina. 2016. *Covert modals in root contexts*. Massachusetts Institute of Technology Doctoral dissertation.

Oikonomou, Despina. 2022. Detecting variable force in imperatives: A modalized minimal approach. *Natural Language and Linguistic Theory*, 1–56. DOI: https://doi.org/10.1007/s11049-022-09554-1

Petrone, Caterina & D'Imperio, Mariapaola. 2011. From tones to tunes: Effects of the f 0 prenuclear region in the perception of Neapolitan statements and questions. In Frotz, Sonia & Elordieta, Gorka & Prieto, Pilar (eds.), *Prosodic categories: Production, perception and comprehension,* 207–230. Berlin, Germany: Springer Verlag. DOI: https://doi.org/10.1007/978-94-007-0137-3_9

Pierrehumbert, Janet & Hirschberg, Julia. 1990. The meaning of intonational contours in the interpretation of discourse. In Cohen, Phillip & Morgan, Jerry & Pollack, Martha (eds.), *Intentions in communication*, 271–312. Cambridge, MA: MIT Press.

Portner, Paul. 2004. The semantics of imperatives within a theory of clause types. In *Semantics and Linguistic Theory* 14, 235–252. DOI: https://doi.org/10.3765/salt.v14i0.2907

Portner, Paul. 2007. Imperatives and modals. *Natural Language Semantics* 15(4). 351–383. DOI: https://doi.org/10.1007/s11050-007-9022-y

Portner, Paul. 2018. Commitment to priorities. In Fogal, Daniel & Harris, Daniel W. & Moss, Matt (eds.), *New work on speech acts,* 296–316. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198738831.003.0011

Ritchart, Amanda & Arvaniti, Amalia. 2013, December. The use of high rise terminals in Southern Californian English. In *Proceedings of Meetings on Acoustics 166 ASA* 20(1), p. 060001. Acoustical Society of America. DOI: https://doi.org/10.1121/1.4863274

Roberts, Craige. 2015. Conditional plans and imperatives: A semantics and pragmatics for imperative mood. In *Proceedings of the 20th Amsterdam colloquium,* 353–362.

Rudin, Deniz. 2018. *Rising above commitment*. University of California, Santa Cruz Doctoral dissertation.

Schwager, Magdalena. 2006. Interpreting Imperatives. Johann Wolfgang Goethe Universität Doctoral dissertation.

Shue, Yen-Liang & Keating, Patricia & Vicenik, Chad & Yu, Kristine. 2011. VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences,* 1846–1849. Hong Kong. Program available online at http://www.seas.ucla.edu/spapl/voicesauce/.

Silverman, Kim & Beckman, Mary & Pitrelli, John & Ostendorf, Mari & Wightman, Colin & Price, Patti & Pierrehumbert, Janet & Hirschberg, Julia. 1992 TOBI: a standard for labeling English prosody. *Proc. 2nd International Conference on Spoken Language Processing* (ICSLP 1992), 867–870. DOI: https://doi.org/10.21437/ICSLP.1992-260

Starr, William. 2020. A preference semantics for imperatives. *Semantics and Pragmatics* 13(6). 1–62. DOI: https://doi.org/10.3765/sp.13.6

Starr, William B. 2010. Conditionals, meaning and mood. New Brunswick, NJ: Rutgers University Ph.D. thesis.

Steedman, Mark. 1991. Structure and intonation. *Language.* 260–296. DOI: https://doi.org/10.1353/lan.1991.0098

Steedman, Mark. 2007. Information-structural semantics for English intonation. In Lee, Chungmin & Gordon, Matthew & Büring, Daniel (eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation* (Studies in Linguistics and Philosophy 82), 245–264. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-1-4020-4796-1_13

Steedman, Mark. 2014. The surface-compositional semantics of English intonation. *Language.* 2–57. DOI: https://doi.org/10.1353/lan.2014.0010

Surányi, Balázs & Turi, Gergő. 2020. Intonational effects on English scopally ambiguous sentences. *Ilha do Desterro* 73(3). 13–36. Florianópolis. DOI: https://doi.org/10.5007/2175-8026.2020v73n3p13

Truckenbrodt, Hubert. 2012. Semantics of intonation. In Maienborn, Claudia & von Heusinger, Klaus & Portner, Paul. (eds.), *Semantics: An International Handbook of Natural Language Meaning, 2039–2069.* Berlin: Mouton de Gruyter.

Truckenbrodt, Hubert. 2020. Semantics of English intonation: "A leopard? A leopard!". In Gutzmann, Daniel & Matthewson, Lisa & Meier, Cécile & Rullmann, Hotze & Zimmermann, Thomas E. (eds.), *The Wiley Blackwell Companion to Semantics,* 1–26. Wiley. DOI: https://doi.org/10.1002/9781118788516.sem025

von Fintel, Kai & Iatridou, Sabine. 2017. A modest proposal for the meaning of imperatives. In Arregui, Ana & Rivero, Maria L. & Salanova, Andrés P. (eds.), *Modality across syntactic categories,* 288–319. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780198718208.003.0013

Weirich, Melanie & Simpson, Adrian. P. 2014. Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics* 43. 1–10. DOI: https://doi.org/10.1016/j.wocn.2014.01.001

Xu, Yi & Xu, Ching. X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33(2). 159–197. DOI: https://doi.org/10.1016/j.wocn.2004.11.001