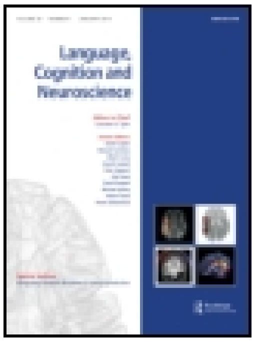

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Mapping prosody onto meaning – the case of information structure in American English

Timo B. Roettger, Tim Mahrt & Jennifer Cole

To cite this article: Timo B. Roettger, Tim Mahrt & Jennifer Cole (2019): Mapping prosody onto meaning – the case of information structure in American English, Language, Cognition and Neuroscience

To link to this article: https://doi.org/10.1080/23273798.2019.1587482

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2019.1587482

Routledge
Taylor & Francis Group

REGULAR ARTICLE

Mapping prosody onto meaning – the case of information structure in American English*

Timo B. Roettger[a], Tim Mahrt[b] and Jennifer Cole[a]

[a]Department of Linguistics, Northwestern University, Evanston, IL, USA; [b]Wovn Technologies, Inc., Tokyo, Japan

ABSTRACT

Prosody is a central part of human speech, with prosodic modulations of the signal expressing important communicative functions. Yet, the exact mechanisms of how listeners map prosodic aspects of the speech signal onto speaker-intended discourse functions are only poorly understood. Here we present three perception experiments that test the mapping between the prosodic form of a heard utterance and possible information structural categories (here: focus and givenness) determined by a discourse context. Results suggest varying degrees of accuracy dependent on the specific information structure categories that are presented to the listener in the experiment (the target and the competitor). Moreover, listeners are sometimes biased towards or against certain discourse contexts. These biases are compatible with the idea that listeners infer speaker intentions based not only on bottom-up processing of acoustic cues but also on probabilistic knowledge about how likely prosodic forms co-occur with specific discourse contexts.

ARTICLE HISTORY
Received 8 July 2018
Accepted 12 February 2019

KEYWORDS
Prosody; intonation; focus; information structure; American English

1. Introduction

The prosodic form of a linguistic expression is an integral part of signalling meaning in human language. Prosody can not only encode emotions, speaker involvement, and attitude, it also plays a crucial role in expressing linguistic meaning: It conveys the intended illocutionary act, structures the utterance into smaller meaningful units, and allows the speaker to emphasise certain units while deemphasising less important information. Given the importance of all of these dimensions of meaning for successful communication, our knowledge about how prosody guides listeners' interpretation of utterance meaning is surprisingly limited.

A central concern for a theory of prosodic meaning is how intonational form maps onto discourse functions. For example, information structure (the division of sentences into focus and background) and information status (the degree of activation of a referent in the current discourse model) can be expressed by certain prosodic parameters. Some authors have proposed a direct mapping of acoustic parameters onto information structural categories (e.g. Cooper, Eady, & Mueller, 1985; Fry, 1955), others have proposed that phonological categories mediate acoustics and discourse functions (e.g. Ladd, 2008; Pierrehumbert, 1980). Regardless of its phonological interpretation, it has been argued that information structure and information status can be expressed through the assignment of phrasal prominence (i.e. positioning the word in a strong metrical position, such as the head of the prosodic phrase) and the association of pitch accents (i.e. tonal events co-occurring with lexically stressed syllables) in English (e.g. Brown, 1983; Büring, 2006; Chafe, 1987; Ladd, 2008; Rooth, 1992; Selkirk, 1995). Different pitch accents have been described to express different types of discourse relations. For instance, a pitch accent with a late (and high) fundamental frequency (f_0) peak and a rising onglide (L + H* in the ToBI annotation) is described as signalling contrastive focus; A pitch accent with a medial peak and shallow rising onglide (H*) is described as signalling new information (cf. Pierrehumbert & Hirschberg, 1990; Watson, Tanenhaus, & Gunlogson, 2008).

A challenge for theories of prosodic meaning is seen in detailed empirical studies on several languages showing that implicitly assumed one-to-one-mappings between pitch contours and discourse function do not hold for all speakers of a language, or even for one speaker all of the time (German: Cangemi, Krüger, & Grice, 2015; Grice, Ritter, Niemann, & Roettger, 2017;

CONTACT Timo B. Roettger timo.b.roettger@gmail.com

*TM and JC conceived the idea for this research and developed the experimental design. TM generated stimuli, carried out the experiment, and assembled the data. TR analyszed the data statistically and archived materials, data, and scripts in online repositories. TR and JC wrote the manuscript.

Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2019.1587482.

English: Cruttenden, 1986; Peppé, Maxim, & Wells, 2000; Turnbull, 2017; Tashlhiyt: Roettger, 2017). For example, Grice et al. (2017) present evidence from a German speech production experiment. Prompted by discourse setting questions, speakers had to produce utterances with different focus structures (Broad, Narrow, Contrastive, No focus). Some speakers produced different pitch accents for the same focus category and other speakers produced one and the same pitch accent for different focus categories. Similarly, Roettger (2017) shows that speakers of Tashlhiyt Berber can prosodically encode questions and contrastive statements with a rise-fall in pitch on the phrase-final word. This tonal event can either occur on the final or prefinal syllable. Both questions and contrastive statements can occur with either final or prefinal rise-falls. However, questions are probabilistically more likely to be produced with a final rise-fall (see also Grice, Ridouane, & Roettger, 2015; Roettger & Grice, 2015).

These studies suggest that there is no one-to-one-mapping between intonational events and speaker intentions; any assumed mapping is probabilistic at best (systematic but not deterministic). More recent work takes such variability into account and provides information as to the statistical distribution of alternative realisations of a given function (e.g. Yoon, 2010 for English; Grice et al., 2017; Baumann, 2006, and Baumann, Röhr, & Grice, 2015 for German; Cangemi & Grice, 2016 for Italian).

Despite this large amount of variability, psycholinguistic work has shown that in some contexts listeners can rapidly anticipate speaker intentions based on intonational information even before disambiguating lexical material is heard (e.g. Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008; Roettger & Franke, 2018a, 2018b; Roettger & Stoeber, 2017; Watson et al., 2008; Weber, Braun, & Crocker, 2006). These studies have demonstrated that listeners show anticipatory eye movements (or hand movements) when hearing an intonational event that allows them to predict an upcoming word based on its status as, e.g. new or given relative to the prior discourse context. This predictive behaviour is not only informed by bottom-up acoustic cues but also by dynamically adaptable probabilistic expectations about likely intonational contours in a given context (Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Roettger & Franke, 2018a, 2018b).

The latter findings are in line with a rational analysis approach (Anderson, 1990) to speech perception (e.g. Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015; Kleinschmidt, Weatherholtz, & Florian Jaeger, 2018; Norris, McQueen, & Cutler, 2003), assuming that speech perception can be thought of as a process of *inference under uncertainty*: listeners know that each linguistic unit is realised as a distribution of acoustic cues. The listener probabilistically infers how likely each possible linguistic unit is, considering their knowledge of these cue distributions within a given context. This inference process is informed by many different information sources, including information associated with the speaker and the discourse context. Prosodic processing as inference under uncertainty can account for successful perception of prosodic information despite its ubiquitous variability. It simultaneously allows for the integration of top-down information. This account contrasts with traditional models of perception of prosody that implicitly or explicitly assume a simple mapping of acoustic cues onto respective discourse functions.

Taking the systematic but probabilistic nature of mapping prosodic form onto discourse function into account, listeners should in principle have some ability to distinguish discourse functions based on only prosodic information, even in contextually impoverished contexts (e.g. in a controlled experiment). At the same time, listeners' performance should be poor when their task is devoid of communicative context and when they are not able to adapt to a given situation, because expectations from prior discourse are impoverished or missing, decreasing the influence of top-down processing on perceiving prosody.

The present paper tests to what extent prosodic form-function relationships can be detected on the basis of prosodic cues. To that end, we test how well listeners detect and distinguish prosodic forms expressing different types of information structural relations: Givenness and Focus distinctions, which have been prominently discussed in the literature as important discourse functions expressed by prosody, most notably, in West Germanic languages (Büring, 2006; Ladd, 2008; Rooth, 1992; Selkirk, 1995). We define focus here according to an alternative semantics account as proposed by Rooth (1992). Focus is a semantic attribute of a word or phrase signalling that the proposition or parts of it have discourse-relevant alternatives. Focus can differ with respect to the location and scope of its domain.

Focus types can be marked by morphosyntactic devices such as word order or focus particles. Alternatively, in English and German, focus is often described as being signalled only by intonation with the position and type of pitch accents differentiating between focus type and scope. Acoustic correlates of focus and information status distinctions have been identified from experimental and corpus studies of English. In English, the nuclear prominence is located by default on the rightmost (content) word in the prosodic phrase (Chafe, 1987; Pierrehumbert, 1980; Selkirk, 1995). Nuclear prominence can be assigned to a word in an earlier position in the phrase if that word is focused and if the phrase-final word is lexically or referentially given. Speakers often distinguish a focus-marking prominence from a non-focus-marking prominence through scaling and alignment of the pitch contour (Breen, Fedorenko, Wagner, & Gibson, 2010). Such differences are analysed by some authors as differences in the tonal specification of the pitch accent, with high rising pitch accents (L + H* within the ToBI annotation) being the preferred pitch accent for focused words (Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980; Pierrehumbert & Hirschberg, 1990), while others consider the scaling and alignment differences as gradual in nature (Calhoun, 2006, 2012; Ladd & Schepman, 2003). Given that focus and information status distinctions are reflected in production in the form of measurable differences in acoustic parameters, there is a basis for experimental hypotheses that listeners use the same acoustic parameters as cues to recover focus and information status of words in comprehending speech.

There are several empirical studies that have investigated the perceptual detectability of prosodic focus marking: Gussenhoven (1983) asked listeners to determine whether the question and answer of a question-answer pair came from the same or a different conversation. He compared broad to narrow focus and reports that at least for certain structures there is a perceptible difference between narrow and broad focus, but listeners cannot use this information to reliably tell in which context the sentence was uttered, suggesting that listeners cannot easily associate focus types with respective acoustic forms.

In Welby (2003), English listeners rated a sentence like “I read the DISPATCH” with a pitch accent on “dispatch” as similarly acceptable to questions with either narrow focus (i.e. “What newspaper do you read?”), or broad focus (i.e. “How do you keep up with the news?”), suggesting that listeners cannot easily tease focus types apart based on the acoustic form of the utterance only.

In Rump and Collier (1996), Dutch listeners judged which of four focus structures (neutral, double focus, focus on subject, focus on object) was most likely signalled by resynthesised intonation contours. Listeners were not consistent with respect to how they matched contour and focus structure and some pitch contours remained ambiguous with respect to focus. Other contours were more consistently classified as signalling a particular focus structure.

Breen et al. (2010) asked English listeners to match a recorded statement presented auditorily to a question that sets the discourse context for the statement. Their results indicate that listeners were generally accurate in identifying the focus position (subject focus, verb focus, object focus), but were often not able to differentiate different types of focus on the same constituent. In their experiment, listeners had to choose between seven different response options, making the task particularly challenging.

Cangemi et al. (2015) asked German listeners to identify four different focus types. Stimuli were taken from a production corpus, in which five speakers produced utterances with different focus conditions (broad, narrow, contrastive, no focus) on the same sentential argument, where each focus condition was prompted by a preceding question. In the perception study, listeners heard these sentences and had to select in a four-alternative forced choice task, which among the four prompting questions provided an appropriate discourse context for the heard sentence. They report on categorisation accuracy above chance performance for all focus categories. Their experimental design, however, allowed for an exceptional high degree of accommodation to the stimuli: Speaker productions occurred in separate blocks, i.e. speakers were not interspersed with each other, giving listeners ample opportunity to “tune” into speaker idiosyncrasies. Moreover, the speech material was segmentally very homogenous. Utterances only differed with respect to the quality of the stressed vowel of the target noun (Bieber, Bahber, Bohber), calling listeners’ attention to prosodic differences expressed in that region. Nevertheless, this study provides evidence that German listeners can detect focus types based on prosodic form, at least in some conditions.

All in all, the literature on intonation-based focus perception is characterised by a wide variety of methodologies employed. Studies mainly differ in the type of task (acceptability judgements: Welby, 2003; naturalness judgement: Gussenhoven, 1983; or question-answer congruence: Breen et al., 2010; Cangemi et al., 2015; Rump & Collier, 1996). The latter studies utilising question-answer matching tasks differed also with respect to the number of response options (four response alternatives in Rump & Collier, 1996 and Cangemi et al., 2015 and seven in Breen et al., 2010). The results of these studies reveal an empirically mixed picture and its methodological diversity makes accumulation of evidence difficult. With the exception of Cangemi et al. (2015) on German, none of the above studies was able to clearly show that listeners can detect focus type based on prosodic information. For American English in particular, there is no compelling evidence to date that listeners perceive a difference between focus types such as broad, narrow, and contrastive focus. Whether listeners can use prosody to recognise speaker-intended focus structures remains, however, an important empirical question: Given the inherent probabilistic nature of mapping prosodic form onto communicative function, it is important to test if listeners can make use of prosodic cues to speaker intentions and if so to what extent they use these cues.

The present study is an effort to provide empirical evidence for the relationship between the prosodic signal and information structure as perceived by a listener. Experimental results are presented to investigate the prosodic form-function mapping in perception by asking (i) how well listeners can identify the focus condition of an utterance based on its prosodic form (form-to-function mapping), and (ii) how well they can identify an appropriate prosodic form to match the focus condition specified by the discourse context (function-to-form mapping). Similar to several prior studies, the present study uses question-answer congruence, which provides a detailed view of the form-function mapping perceived by listeners. However, our studies differ from prior studies in reducing the complexity of the experimental task, towards the goal of minimising task effects on the listeners' judgments of form-function association.

2. Methods

This paper presents a series of experiments exploring listeners' perception of the relationship between the prosodic form of an utterance and its focus conditions as established from the immediate discourse context.[1] Here we describe the methodology and statistical analysis employed. Section 2.1 presents the focus categories tested as they relate to theories of information structure and information status; Section 2.2 describes the experimental stimuli; Section 2.3 describes the design and procedures; Section 2.4 discusses the statistical methods we use to test our hypotheses.

2.1. Information structure categories

In the following experiments, listeners reacted to short question-answer dialogues in which the question provides the discourse context that establishes one of four information structure conditions of the answer: broad focus, narrow focus, and contrastive focus on the sentential subject, and the sentence subject as discourse-given. We adopt the question-answer congruence paradigm and operationalise focus following Büring (2012): in an answer, focus marks that constituent which can be construed as corresponding to a wh-phrase in a preceding question. Consider the following example.

(1) Damon fried the omelet.

a. Do you know what happened yesterday?	[Damon fried the omelet]$_F$.
b. Do you know who fried the omelet?	[Damon]$_F$ fried the omelet.
c. Do you know what Damon fried?	Damon fried [the omelet]$_F$.
d. Did Pam fry the omelet?	[Damon]$_F$ fried the omelet.
e. Did Damon fry the omelet?	Damon fried the omelet.

The statement in (1) is a suitable morphosyntactic construction to answer all questions in (a-e), they only differ in their focus structure. Question (a) elicits whole-sentence focus (also referred to as "broad focus"). A sentence has broad focus when it is uttered in an out-of-the-blue context, in the absence of a preceding discourse context, or with no particular correspondence to a preceding context. For a sentence with broad focus, the entire proposition expressed by the sentence is in focus and all constituents constitute new information. A common question used to elicit broad focus is "What happened?" or "What is new?" (e.g. question-answer pair 1a).

Questions (b-c) elicit "narrow focus" either on the subject (b) or the object (c). A narrow focus sentence is one that contains a constituent that introduces relevant, new information to the discourse. The constituent with narrow focus may provide the answer to a wh-question, or it may highlight new information that is relevant to the discourse context, e.g. as an elaboration of information already given. In example (b) "Damon" contrasts with an open set of alternatives to the subject (all entities that could have fried an omelet), while in (c) "the omelet" contrasts with an open set of alternatives to the thematic object (all possible things Damon could have fried).

In (d), the focused constituent is explicitly contrasted with the alternative in the question ("Pam"), and constitutes a specific type of narrow focus, which is referred to as "contrastive focus" (or "corrective focus"). Similar to a narrow focus sentence, a sentence with contrastive focus contains a constituent (here "Damon") that relates specifically to an element of the preceding discourse (here "Pam"). Contrastive focus marks the referent of the constituent as singled out from a set of possible alternatives made salient by the discourse context (Rooth, 1992).

A sentence that cannot be construed as providing an answer to a wh-question or as specifying a contrastive referent may lack focus altogether. Such an example is illustrated in (1e), where all elements in the sentence are discourse-given, both lexically (the words are explicitly mentioned in the preceding question) and referentially (the referent of each word and phrase is established in the preceding discourse). We refer to such sentences as "given" in what follows.

2.2. Stimuli

The stimuli used in the perception experiments were selected from productions of nine different English sentences (see 2). Each sentence was produced four times, once for each of the four focus categories described in the preceding section (Broad, Narrow, Contrastive, Given). The stimuli were recorded in a soundproof booth with a high-quality, head-mounted microphone. One informed female speaker of American English produced all of the stimuli (see osf.io/4qxmh/). To make her productions as natural as possible, the sentences were produced in a live dialogue enacted with the experimenter who asked questions (see 1) that prompted the speaker to produce an appropriate full sentence response for each of the four focus conditions described in Section 2.1.

(2) (a) Daisy warned the owner.
(b) Damon fried the omelet.
(c) Dorah filmed the movie.
(d) Harry raised the window.
(e) Jamie dyed the laundry.
(f) Jonny helped the warden.
(g) Jonah burned the onion.
(h) Maddie found the TV.
(i) Mary rolled the barrel.

Except for the discourse-given context, the full sentence responses were not read aloud from text but were formulated by the speaker as appropriate full sentence responses to the experimenter's question (for a full list, see Appendix 1), inserting in subject position a name that was presented in written form for each sentence trial. The full sentence responses for the given condition (no focus) were written out and read aloud by the speaker. This was done to avoid the inadvertent production of a sentence with pronominal elements (e.g. "Yes, she warned him" for 2a), which are also acceptable responses to a polar question (e.g. "Did Daisy warn the owner?" for 2a). The recorded sentences were normalised for amplitude based on the peak amplitude of the entire recording session, using the normalise function in Audacity (Audacity Team, 2015) with the DC offset removed and peak amplitude normalised to −1.0 db. This process resulted in 36 utterances (9 utterances for 4 different focus categories) of roughly equal amplitude.

Auditory inspection of stimuli by two native speakers (TM and JC) determined that the focus categories were produced with intonation patterns that sounded natural and congruent with the matched discourse prompt (i.e. the context question). Qualitative characterisation of f_0 contours, based on ToBI criteria, reveals that the answers exhibit expected intonational contours, with distinct contours for each of the four focus conditions. Figure 1 shows time-normalised f_0 contours and ToBI labels for each of the 9 sentences (grey) alongside the mean contour (in colour), with productions grouped by the intended focus condition (i.e. focus conditions determined by the question prompt for each production).[2] The differences among the four focus categories can be seen in the f_0 contours on the subject and object positions. In the subject position, the broad and contrastive focus conditions show a noticeable rise-fall contour, the narrow focus condition has a shallower rise-fall (and in some tokens just a shallow fall), and the given condition exhibits a relatively low f_0 with an even shallower fall. In the object position, the given and narrow focus conditions show a flat or mildly falling f_0 excursion that extends with a nearly even slope across the interval of the object noun. The broad focus condition shows a noticeable rise-fall f_0 contour on the object, while the contrastive focus condition exhibits a low plateau that ends in a sharp fall (or in one instance, a rise) to the end of the utterance.

Figure 2 shows the raw values for the maximum f_0 values of the sentence subject and the sentence object. The f_0 max values for the subject overlap substantially between the broad and contrastive focus conditions, and between the narrow and given conditions. While the overlap between broad and contrastive is resolved when looking at f_0 maxima on the sentence object (clear separation), given and narrow remain highly overlapping. We will come back to these different degrees of overlap later.

To establish that the stimuli were acoustically differentiable into four prosodically distinct classes, we submitted the stimuli to linear discriminate analysis (LDA). We first inspected a variety of measures of pitch, intensity, and duration that were extracted from the subject, verb, and object positions of the sentence utterances. These were fed into an LDA analysis in the R *MASS* package (Venables & Ripley, 2002). The function indicated that there was collinearity among some of the measures. To determine which variables exhibited collinearity, the correlation of all possible feature pairs was taken and, for each correlated pair, only one of the features was chosen to be in the final set of acoustic features. With the final selection of acoustic features chosen, the LDA was run again. Using a leave-one-out (LOO) analysis, the LDA was able to discriminate the four focus categories with high accuracy (91.4%).

Note that we do not take the LDA results as proxy for human judgments of perceptual distinctiveness. Rather, the LDA analysis serves to independently verify that there is a basis for perceptual distinction in measurable acoustic distinctions that are sufficient for classification by statistical methods. It is important to note that the LDA analysis indicates some degree of overlap

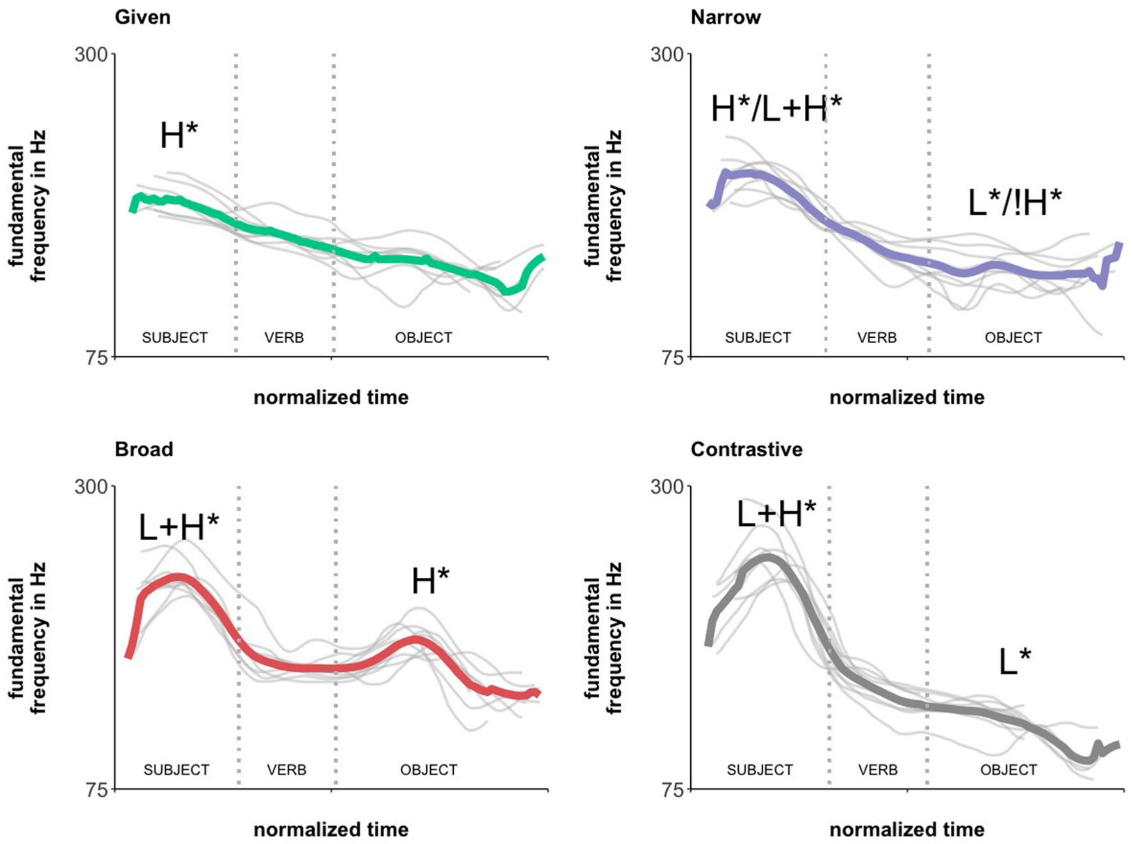

Figure 1. Smoothed and interpolated f_0 contours of the acoustic stimuli (grey) alongside the average f_0 contour (coloured) for all four focus conditions.

between the four focus categories, which suggests a degree of acoustic ambiguity in some of the productions for the acoustic measures. It is possible that using other acoustic measures, the distinction between the four focus categories might have been better captured, leading to higher accuracy. Our concern here is not primarily about how the four focus categories are differentiated from one another by the stimulus speaker, but rather to demonstrate that the four focus categories are acoustically differentiated.

2.3. Study design

The study was conducted to evaluate listeners' perception of four focus categories in relation to the prosodic form of an utterance. The perception test was operationalised through two tasks: one where listeners had to select which of two prosodic patterns best signalled a specified focus category (1 context – 2 prosodic forms, henceforth 1C-2P); and the other where listeners had to select which of two focus categories was signalled by the prosody of an utterance (2 contexts – 1 prosodic form, henceforth 2C-1P). Rather than presenting participants with all four focus categories in a single (lengthy) experiment, a between-subjects design was chosen, exposing individual participants to only one pair of focus conditions (broad focus vs. given, broad vs. contrastive focus, broad vs. narrow focus, given vs. contrastive focus, given vs. narrow focus, and contrastive vs. narrow focus). A further distinction in the response option given to the participant was introduced: One group of participants used a two-alternative forced-choice response, while another group of participants used a 5-point scale response. In what follows, these experimental conditions are grouped into 3 experiments, Experiments 1 and 2 use the two-alternative forced choice response option, and Experiment 3 uses the 5-point scale response option.

2.3.1. Tasks

In each experiment, auditory stimuli in the form of mini question-answer dialogues were presented to participants. There were two such dialogues on each trial, which differed in the prosodic congruence of the question and answer. Participants were presented with two play buttons on opposite sides of the screen, one for each mini dialogue (Q-A pairing). Participants were allowed to listen to the audio files as many times as

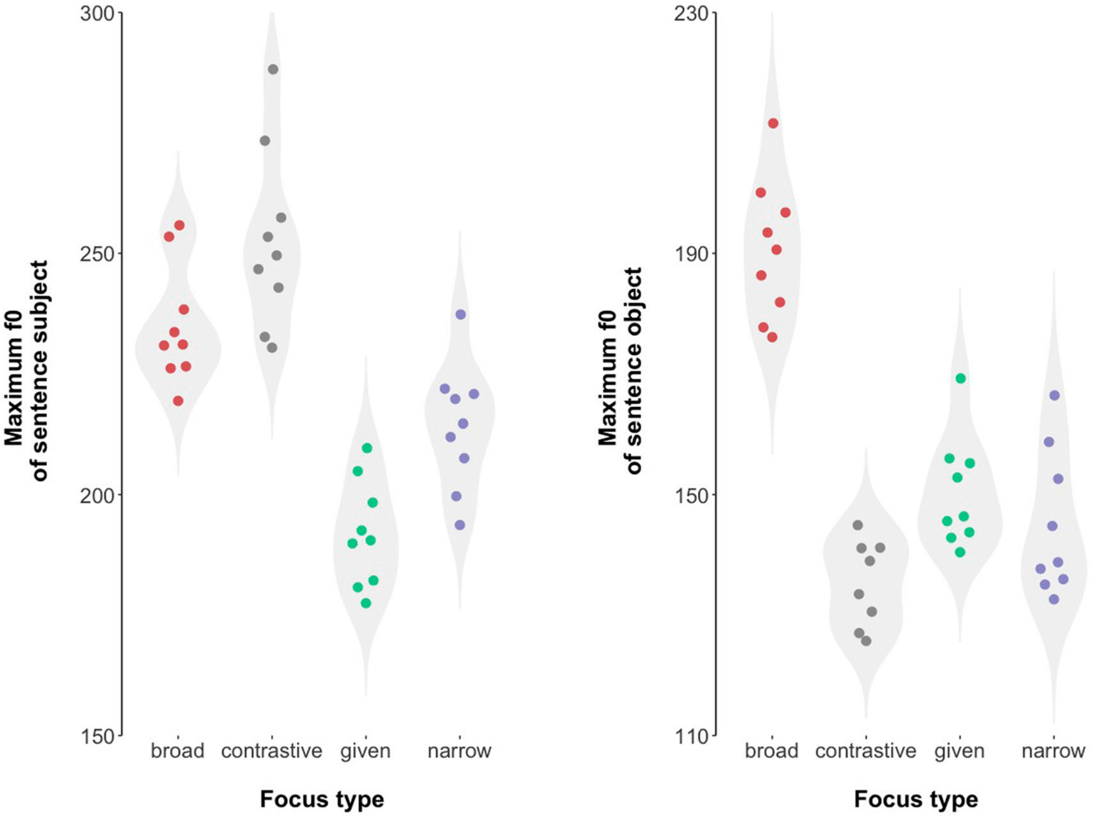

Figure 2. Raw f_0 maximum values of the sentence subject (left panel) and sentence object (right panel) for all target sentences. The grey shapes in the background indicate kernel density curves of these raw values in order to allow for a better visual assessment of overlap between categories.

they wanted before responding. Participants proceeded through the experiment at a self-selected pace. In one dialogue, the question-answer pair was matched in their focus condition (i.e. the answer was produced by the model speaker as a response to the question appearing in the dialogue), while in the other dialogue the question-answer pair was mismatched (the answer was produced by the model speaker in response to a different question than the one appearing in the dialogue). Participants were instructed to either choose the dialogue that sounded the most appropriate or natural (Experiment 1 and 2) or to use a 5-point scale to indicate which of the two dialogues they prefer.

The 1C-2P task tests the mapping from discourse function to prosodic form. In this task, the two dialogues in a trial had the same question, but the question was paired with answers that were prosodically distinct, i.e. from two different categories shown in Figure 1. This task examined whether listeners could identify a preferred acoustic prosodic signal for the particular focus condition specified by the discourse context. Note that the proposition of the answers was always the same for both dialogues and was textually appropriate as a response to the prompting questions. An example of the 1C-2P task is shown in (3), contrasting broad and narrow focus conditions. If narrow focus is prosodically encoded and perceptually detectable by the listener, the dialogue in (3a) with narrow focus prosody (as in Figure 1 above) should sound more natural than the dialogue in (3b) with broad focus prosody (as in Figure 1 above).

(3) Dialogue pair from the 1C-2P task

a. *Incongruous*	
Q: Do you know who ripped the ledger?	[Narrow focus prompt]
A: Yes, [Mary ripped the ledger]$_F$.	[Broad focus prosody]
b. *Congruous*	
Q: Do you know who ripped the ledger?	[Narrow focus prompt]
A: Yes, [Mary]$_F$ ripped the ledger.	[Narrow focus prosody]

The 2C-1P task tests the mapping from prosodic form to discourse function. In this task, the two dialogues had textually different questions, each of which set up different focus conditions for the answer, while the answers were the same in both dialogues (i.e. the same audio file). This task examined whether listeners could identify the discourse context that matched the focus condition of the answer, perceived on the basis of its prosodic form. An example of the 2C-1P task is shown in (4), contrasting broad and narrow focus.

(4) Dialogue pair from the 2C-1P task
a. *Incongruous*
Q: Do you know what happened yesterday? [Broad focus prompt]
A: Yes, [Mary]$_F$ ripped the ledger. [Narrow focus prosody]
b. *Congruous*
Q: Do you know who ripped the ledger? [Narrow focus prompt]
A: Yes, [Mary]$_F$ ripped the ledger. [Narrow focus prosody]

The two experimental tasks (1C-2P, 2C-1P) were designed to explore possible sources of ambiguity stemming from overlap in the range of acoustic patterns that a speaker may produce in a certain focus condition, and which a listener may judge as acceptable acoustic cues for perceiving distinctions in focus-related meaning. If listeners can successfully detect prosody-focus mappings, then participants in the 2C-1P and 1C-2P tasks should be equally accurate in identifying the most natural sounding dialogue in each trial. If participants perform poorly in the 1C-2P task (choosing from two prosodically distinct answers), that would suggest that a range of acoustic cues can signal the same meaning. If participants perform poorly in the 2C-1P task (choosing from two textually distinct context questions), it would suggest that a certain acoustically specified prosodic pattern may be congruent with multiple focus-related meanings (as specified by the discourse context).

The same dialogues with the same acoustic stimuli were used for all three experiments. In experiment 1, participants performed only the 1C-2P task in a forced choice design. In experiment 2, participants performed only the 2C-1P task in a forced choice design. Experiment 3 comprised both 1C-2P and 2C-1P tasks but with more nuanced response options on a 5-point Likert scale:

- *only left*: only the dialogue on the left side of the screen sounded natural
- *left preferred:* both dialogs sound natural, but the left dialogue is preferred
- *equally good:* both dialogs sound equally natural and acceptable
- *right preferred*: both dialogs sound natural, but the right dialogue is preferred
- *only right*: only the dialogue on the right side of the screen sounded natural

Only in Experiment 3, 1C-2P and 2C-1P trials were presented in different trials to the same participants.[3] Participants in Experiment 3 were instructed that sometimes the answers would vary in the way they were said and sometimes the questions would vary. None of the experiments gave participants any explicit instructions or training regarding the information structure or prosody of the dialogs they would hear.

Experiment 1 and 2 consisted of 18 trials presenting a pair of Q-A dialogues for one of the six different focus condition pairs (broad-given, broad-contrastive, broad-narrow, given-contrastive, given-narrow, contrastive-narrow). Each of the 9 stimulus sentences in (2) was presented as the answer in two trials that differed in which of the two focus categories was specified in the congruent dialogue. For example, the dialogue pairs in (3) and (4) are taken from the group testing broad vs. narrow focus prosody. In the trials shown in (3) and (4), it is the (b) dialogues that are congruent – the focus condition prompted by the question matches the focus condition that is expressed by the prosody of the answer. These same experiments (testing the broad vs. narrow focus categories) included another trial with "Mary ripped the ledger" in which the congruent dialogue matches the question and answer in the broad focus condition. The order of the stimuli was pseudorandomised, i.e. shuffled by hand such that no two consecutive items contained the same lexical content (i.e. the same answer sentence).

Experiment 3 consisted of 36 trials, presenting a pair of Q-A dialogues for one of six different focus condition pairs for both 2C-1P and 1C-2P tasks.

2.3.2. Participants

Participants were recruited online via the crowd-sourcing website Amazon Mechanical Turk (AMT). Participants were restricted to people from the United States via filtering of IP address by the AMT system, and were restricted to be at least 18 years old. The participants came from all around the U.S. Participants who reported themselves as being non-native speakers or indicated that they were born and grew up outside of the U.S. were excluded from the study. To prevent incentivising dishonesty, they were not told that they had to be a native speaker of American English to participate and they were still compensated for their time, regardless of whether or not their data was used. Data that was excluded due to these circumstances was replenished by running additional participants. Data analysis was not initiated before the complete data set was available. Participants were only allowed to participate in one of the three experiments. These constraints were managed by LMEDS (Mahrt, 2016), the web platform used to run all of the experiments.

Experiments 1, 2, and 3 used data from 180 participants each, for a total of 540 participants. Participants in each Experiment (1–3) were randomly assigned to one of six groups testing different pairs of focus conditions. Experiments 1 and 2 took about 15 min to complete, while Experiment 3 took about 25 min. Participants were compensated at a rate of $10/hour.

2.4. Statistical analysis

We submitted participants' responses to Bayesian hierarchical models using R (R Core Team, 2018) and the brms package (Bürkner, 2016). We operate within the Bayesian inferential framework (rather than within a frequentist framework) due to two reasons:

First, Bayesian methods allow us to directly answer the primary question: *How plausible is our hypothesis given the data?* We can answer this question by quantifying our uncertainty about the parameters of interest, which frees us from committing to hard cut-off points for statistical significance (such as the arbitrary 0.05 alpha level).

Second, it is easier to flexibly define hierarchical models (also known as mixed effects or multilevel models) in the Bayesian framework than in the frequentist framework. The frequentist linear mixed model standardly used in quantitative linguistics is generally fit with the lme4 package in R. However, the linear mixed effects models for categorical data that also include the maximal random effects structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013; Schielzeth & Forstmeier, 2009) tend not to converge or to give unrealistic estimates of the correlations between random effects (Bates, Kliegl, Vasishth, & Baayen, 2015). Such non-convergence issues are particularly severe for logistic regression models (Kimball, Shantz, Eager, and Roy 2018). In contrast, the maximal random effects structure can be fitted without problems using Bayesian hierarchical models.

We used different statistical models for Experiments 1 and 2 than for Experiment 3. For Experiments 1 and 2 we fit a hierarchical logistic regression model to response accuracy (binomial: correct vs. incorrect) predicted by the target focus category in the congruent dialogue (4 levels: Given, Broad, Narrow, Contrastive), the competitor focus category (3 levels, e.g. for the target category Broad, the competitor focus would be Narrow, Contrastive, or Given) and their two-way interaction. The models included a maximal random-effect structure, including a random intercept for subjects (since it is a between-subject design), and a random slope allowing the predictor interaction to vary by experimental items (the 9 sentences comprising the experimental stimuli).

We used weakly informative Gaussian priors centred around zero with $\sigma = 5$ for all population-level regression coefficients. Four sampling chains with 2000 iterations were run for each model, with a warm-up period of 1000 iterations. We report, for each parameter of interest, 95% credible intervals and the posterior probability that a coefficient parameter β is bigger than zero $Pr(\beta > 0)$. A 95% credible interval demarcates the range of values that comprise 95% of the probability mass of our posterior beliefs, such that no value outside the CI has a higher probability than any point inside of it (see, e.g. Jaynes & Kempthorne, 1976; Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). We judge there to be compelling evidence for an effect if zero is (by a reasonably clear margin) not included in the 95% CI and $Pr(\beta > 0)$ is close to zero or one.

For Experiment 3, where the same participants performed both tasks (1C-2P and 2C-1P) we ran two subset analyses on the data, one that models the 1C-2P trials and one that models the 2C-1P trials. Recall that Experiment 3 used an elaborated set of five response options. For both tasks, we fitted Bayesian hierarchical ordinal logistic models to the ordered response options predicted by the target focus category of the congruent dialogue (4 levels, as for Exps. 1 and 2), the competitor focus category (3 levels, as for Experiments 1 and 2) and their two-way interaction. The five responses were re-labelled as follows: If the congruent question-answer pair was on the right side of the screen, we binned "always right" as "always", "right preferred" as "preferred", "equally good" as "equal", "left preferred" as "dispreferred", and "always left" as "never". Responses were similarly re-labelled for trials in which the congruent question-answer pair was on the left side of the screen, by swapping "left" for "right" in the re-labelling scheme. The re-labelled responses were rank ordered: never > dispreferred > equal > preferred > always. The models for Experiment 3 included a random intercept for subjects (since main effects of focus conditions were tested in a between-subject design), and a by-item random slope allowing the predictor interaction to vary by experimental items. We used weakly informative student-*t* priors centred around zero with $\sigma = 1$ and $dfs = 5$ for all population-level regression coefficients. The inferential criteria are the same as discussed for Experiments 1 and 2.

Posterior probabilities tell us the probability that the parameter has a certain value (given the data and model); note that these probabilities are not frequentist *p*-values. Note also that there is no notion of Type-I or Type-II errors in Bayesian statistics because the inference does not depend on hypothetical repetitions of the experiment; the data are evaluated on their own merits, and no supposition is made about the replicability of the effect. In order to present statistics as close to widely used frequentist practices, we chose to define an inferential criterion that seems familiar (95%), but the strength of evidence should not be taken as having clear cut-off points (such as in a null-hypothesis significance testing framework). In line with standards of reproducible research, the data tables and the scripts for the statistical analyses are publicly available here: http://osf.io/4qxmh.

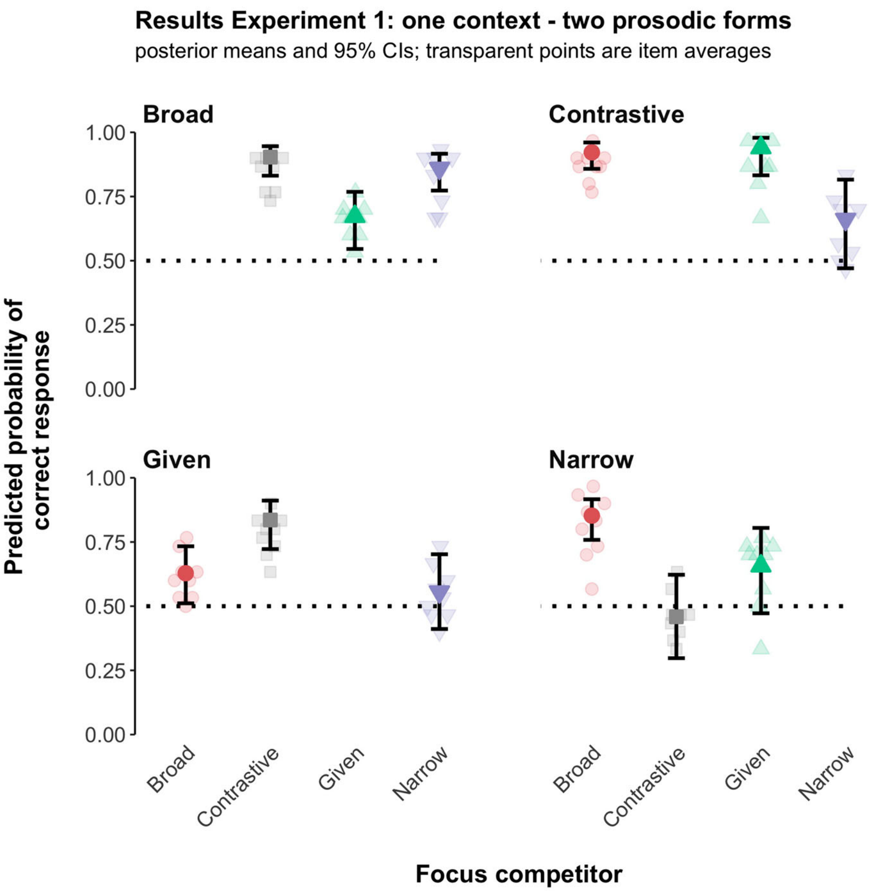

Figure 3. Mean posteriors and 95% credible intervals for the results of Experiment 1, showing predicted accuracy across target focus conditions (in the four panels), and their accompanying focus competitors (x-axis). Semi-transparent small points are average values for each experimental item (sentence). The dotted line indicates chance performance.

3. Results for experiment 1: One context – two prosodic forms

Figure 3 and Table 1 summarise the posterior distribution across conditions for experiment 1. Instead of interpreting regression coefficients, we directly calculate the posterior distribution and accompanying credible intervals for each condition (given the data and the model). We can further directly calculate the probability of respective accuracy estimates being above chance (log odds > 0).

Table 1. Summary of posterior distributions for Experiment 1: Posterior means (95% credible intervals in brackets) for all focus combinations alongside the probability that the estimate is above chance level (log odds > 0) given the data and the model.

Target	Competitor	Estimate	$P(\beta > 0)$
Broad	Contrastive	0.9 (0.83,0.95)	1.00
Broad	Given	0.67 (0.55,0.77)	1.00
Broad	Narrow	0.86 (0.77,0.92)	1.00
Contrastive	Broad	0.92 (0.86,0.96)	1.00
Contrastive	Given	0.94 (0.83,0.98)	1.00
Contrastive	Narrow	0.66 (0.47,0.82)	0.96
Given	Broad	0.63 (0.51,0.73)	0.98
Given	Contrastive	0.84 (0.72,0.91)	1.00
Given	Narrow	0.56 (0.41,0.7)	0.77
Narrow	Broad	0.85 (0.76,0.92)	1.00
Narrow	Contrastive	0.46 (0.3,0.62)	0.31
Narrow	Given	0.66 (0.47,0.8)	0.96

Looking at the estimates, overall, listeners performed well in the task. However, there are obvious interactions between target (henceforth X^T) and competitor categories (henceforth X^C), with varying accuracy estimates for different combinations of categories. Dependent on the competitor category for a trial, listener performance differs tremendously: Except for $Given^T$ competing with $Narrow^C$, and $Narrow^T$ competing with $Contrastive^C$, all conditions show evidence for above chance accuracy. Listeners thus seem to be able to infer the intended prosodic information in the signal based on the discourse setting question.

Listeners' performance differed, however, as a function of which categories were compared. For $Broad^T$ (upper left panel), listeners exhibit higher accuracies when the competitor is $Contrastive^C$ ($\beta = 0.90$ [0.83,0.95]) or $Narrow^C$ ($\beta = 0.86$ [0.77,0.92]) than when the competitor is Given^C ($\beta = 0.67$ [0.55,0.77)]; For Contrastive^T (upper right), listeners exhibit higher accuracies when the competitor is Broad^C ($\beta = 0.92$ [0.86,0.96]) or Given ($\beta = 0.94$ [0.83,0.98]) than when it is Narrow^C ($\beta = 0.66$ [0.47,0.82]); For Given^T (lower left), listeners exhibit higher accuracies when the competitor is Contrastive^C ($\beta = 0.84$ [0.72,0.91) than when it is Narrow^C ($\beta = 0.56$ [0.41,0.70]); For Narrow^T (lower right), listeners exhibit higher accuracies when the competitor is Broad^C ($\beta = 0.85$ [0.76,0.92]) than when it is Contrastive^C ($\beta = 0.46$ [0.30,0.62]).

The results of this experiment suggest that, in general, listeners can use the prosodic cues available in the signal to distinguish between focus types. Some categories are perceived better than others and accuracy is very much dependent on the competing category. Accuracy was highest for the pairs Contrastive and Given as well as Contrastive and Broad. This is not surprising considering the very distinct f_0 patterns in the stimuli (see Figure 1). Given referents were produced with a high pitch accent (H*), the most frequently occurring pitch accent type, and the smallest f_0 excursion on the subject, while contrastive referents were produced with a high rising pitch accent (L + H*), arguably the most prominent pitch accent type, and the greatest magnitude f_0 excursion on the subject. Broad focus utterances exhibited two prominent pitch accents on the subject and the object, a noticeably distinct utterance-wide pattern.

The other focus pairs were not as well distinguished, including Contrastive and Narrow, and Given and Narrow. The observation that the accuracy between Narrow and every other category is low might be attributed to the acoustic form of Narrow focus utterances. Narrow focus stimuli exhibit an intonational form that greatly overlaps with the other categories. For instance, in Narrow focus stimuli, the subject exhibits a rise-fall contour that is variously labelled as H* or L + H*, but the difference between that and the f_0 contour of the subject in Contrastive focus stimuli, all of which are labelled as L + H* can be characterised as a difference in pitch scaling. Likewise, for the Narrow focus stimuli, the f_0 excursion of the less prominent H* of the subject appears to partially overlap with the f_0 excursion of the Broad focus subject in some instances, and with that of the L* pitch accent of stimuli in the Given category.

In sum, some focus categories elicit lower accuracies while others elicit higher accuracies. These differences may be reflections of different degrees of acoustic overlap. However, overall, listeners seem to be able to match the intended focus type of an utterance to its respective discourse setting question above chance level. The 1C-2P task taps into the question of which acoustic form best conveys the focus condition selected by a particular discourse context, while the 2C-1P task taps into which discourse context best matches the focus condition conveyed by a particular acoustic form

4. Results for experiment 2: Two contexts – one prosodic form

Figure 4 and Table 2 summarise the posterior distribution across conditions for experiment 2. Looking at the estimates, the 2C-1P results differ from the results of experiment 1. Overall, listeners' accuracy is not as high as in the 1C-2P task. This effect is mainly driven by two factor levels: Broad^T and Narrow^C. When Broad focus is the target, listeners were systematically *below* chance, i.e. identifying the utterance as indicating the competitor focus category (Contrastive^C: $\beta = 0.22$ [0.14,0.32]; Given^C: $\beta = 0.30$ [0.22,0.41]; Narrow^C: $\beta = 0.17$ [0.12,0.23]). Beyond showing poor performance in identifying Broad^T, listeners consistently picked the wrong response alternative, suggesting a *bias against* Broad focus. Similarly, when Narrow is the competitor, listeners were systematically below chance, i.e. incorrectly identifying the utterance as indicating Narrow focus (Broad^T: $\beta = 0.17$ [0.12,0.23]; Contrastive^T: $\beta = 0.36$ [0.25,0.48]; Given^T: $\beta = 0.35$ [0.21,0.51]). Again, beyond having difficulty in identifying the target category, listeners were zealous in consistently identifying utterances as Narrow, suggesting a *bias towards* Narrow focus.

In addition to these two biases, listeners had difficulties identifying Given^T and Narrow^T when paired with Contrastive^C, although, in both cases, there is weak evidence that listeners perform above chance (Given^T: $\beta = 0.66$ [0.48,0.84]; Narrow^T: $\beta = 0.65$ [0.48,0.80]).

Our results indicate that when having to identify a discourse context on the basis of the focus condition conveyed by the prosodic form, listeners have substantial difficulties. The observed biases against pairing the Broad^T focus prosody with its matched discourse context and in favour of pairing any prosodic form with the Narrow^C focus discourse context suggests that listeners are influenced by aspects of the context other than the prosodic information in the signal. (Note that listeners were clearly able to use acoustic prosodic cues in the 1C-2P task with the same stimuli). The acoustic prosodic expression used in the Narrow focus stimuli in this study are apparently congruent with a variety of information structure contexts, and similarly, any type of prosodic form is judged as congruent with a Broad focus context. As opposed to that, for Given and Contrastive focus contexts, listeners showed a preference for one prosodic form -the congruent one in this experiment.

The results of experiment 1 and 2 suggest both overlap and differentiation in the association of prosodic

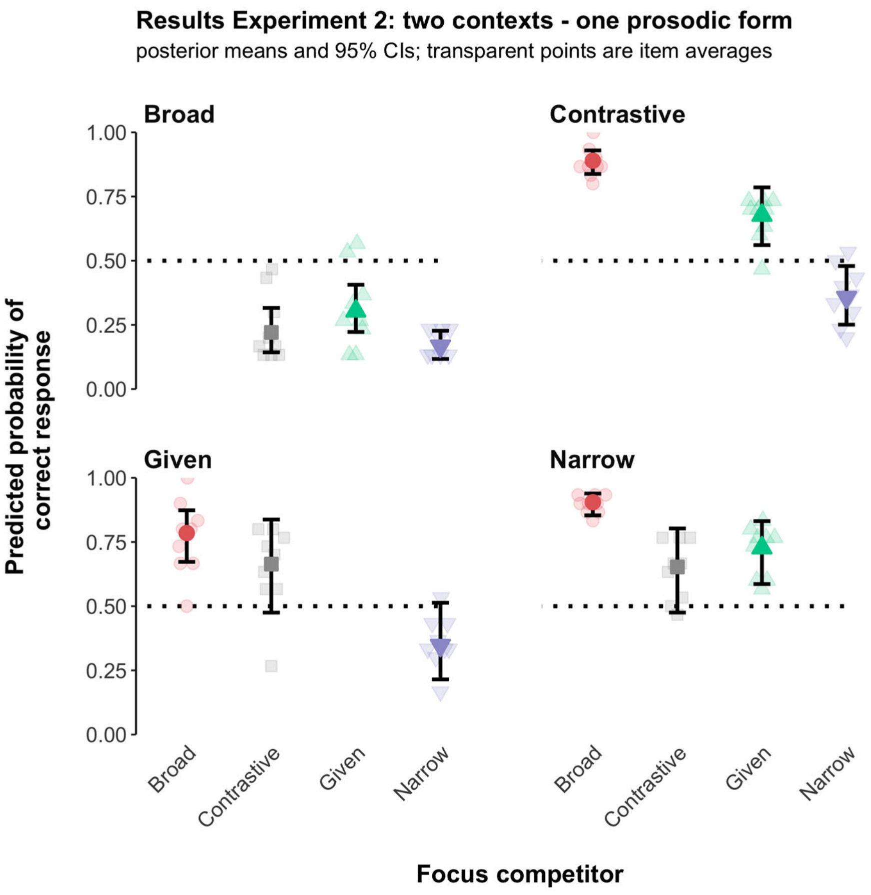

Figure 4. Mean posteriors and 95% credible intervals for the results of Experiment 2, showing predicted accuracy across target focus conditions (in the four panels), and their accompanying focus competitors (x-axis). Semi-transparent small points are descriptive average values for each experimental item (sentence). The dotted line indicates chance performance.

form and focus condition. Differentiation is seen in the finding that listeners show above-chance accuracy in associating prosodic forms with the focus conditions intended by the speaker, for at least some of the distinctions in focus conditions. The associations between form and meaning are far from being one-to-one, and there appear to be ambiguities in both directions of the form-function mapping. This pattern of results may stem from ambiguity in the prosodic encoding of focus that leaves listeners uncertain about the intended focus condition. An alternative account of the results involves listener bias, as suggested in the findings from the 2C-1P task, where given a choice in meaning listeners lean towards or away from inferring certain focus-related meanings.

Table 2. Summary of posterior distribution for Experiment 2: Posterior means (95% credible intervals in brackets) for all focus combinations alongside the probability that the estimate is above chance level (log odds > 0) given the data and the model.

Target	Competitor	Estimate	*Pr*($\beta > 0$)
Broad	Contrastive	0.22 (0.14,0.32)	0.00
Broad	Given	0.3 (0.22,0.41)	0.00
Broad	Narrow	0.17 (0.12,0.23)	0.00
Contrastive	Broad	0.89 (0.84,0.93)	1.00
Contrastive	Given	0.68 (0.56,0.79)	1.00
Contrastive	Narrow	0.36 (0.25,0.48)	0.01
Given	Broad	0.78 (0.67,0.87)	1.00
Given	Contrastive	0.66 (0.48,0.84)	0.95
Given	Narrow	0.35 (0.21,0.51)	0.03
Narrow	Broad	0.91 (0.85,0.94)	1.00
Narrow	Contrastive	0.65 (0.48,0.8)	0.96
Narrow	Given	0.73 (0.59,0.83)	1.00

Experiment 3 seeks to further explore the ambiguity (or bias) in the mapping between prosodic form and focus-related meaning, by offering participants five response options that differ in the strength of association for each of the two form-function mappings presented in each trial.

5. Experiment 3 – Scalar endorsement ratings

The data from Experiment 3 differed from that of Experiments 1 and 2 with respect to available response options. Experiment 3 also differed in presenting participants with 36 trials, 18 per task (9 sentences in two different congruent pairings, as in Experiments 1 and 2). Thus, participants in Experiment 3 produced 18 responses in the same 1C-2P task as those in Experiment 1, and they produced 18 responses in the same 2C-1P task as those in Experiment 2. The data from each task in Experiment 3 was modelled in separate subset analyses, as described in Section 2.5.

5.1. Results for experiment 3: One context – two prosodic forms

Figure 5 and Appendix 2 summarise the posterior distributions across conditions for the 1C-2P task. Overall, participants tend to select the responses "equal", "preferred" and "always" above chance (= 0.2), suggesting that listeners have a general tendency to rate the match between prosodic pattern and focus condition as acceptable, even when the match is incongruent. This is illustrated by the asymmetry of stacked bar plots in Figure 5, which show a greater probability mass in the green bars ("preferred", "always") compared to the red bars ("dispreferred", "never"). (If there was no bias towards either the negative or positive end of the response scale, the stacked bar plots would be symmetrically centred around the horizontal line.)

These general patterns are in line with the results from Experiment 1. Listeners can use the prosodic information in the signal to discriminate intended focus categories above chance levels. However, there is a great amount of variability in how listeners match prosody and focus conditions. Listeners generously endorse utterances as belonging to focus categories other than the one intended by the speaker, indicated here by the large amount of "equal" ratings (both dialogues in the trial rated as equally acceptable). Beyond these general patterns, and in line with our earlier findings, there are also clear differences among responses for different pairings of target focus and competitor focus category.

For Broad$^{\text{T}}$, there is evidence that listeners are more likely to endorse a broad focus prosody correctly paired to a broad focus discourse context when the competitor pairs Contrastive$^{\text{C}}$ prosody with the broad focus context than when the competitor pairing has Given$^{\text{C}}$ prosody. This asymmetry is seen in the comparison of "equal" and "preferred" responses for Broad$^{\text{T}}$ when paired with Contrastive$^{\text{C}}$ ("equal": $\beta = 0.29$ [0.21,0.37]; "preferred": $\beta = 0.58$ [0.52,0.64]) compared to when paired with Given$^{\text{C}}$ ("equal": $\beta = 0.47$ [0.39,0.54]; "preferred": $\beta = 0.4$ [0.3,0.5]) or Narrow$^{\text{C}}$ ("equal": $\beta = 0.45$ [0.37,0.53]; "preferred": $\beta = 0.43$ [0.33,0.53]). This asymmetry in response pattern suggests, again, that listeners find it easier to correctly endorse a Broad$^{\text{T}}$ focus prosody when Contrastive$^{\text{C}}$ focus is the competitor, than with other competitor categories.

In line with that, there is evidence that when paired with Broad$^{\text{C}}$, Contrastive$^{\text{T}}$ elicits fewer "equal" and more "preferred" responses ("equal": $\beta = 0.38$ [0.28,0.48]; "preferred": $\beta = 0.5$ [0.41,0.6]) than when Contrastive$^{\text{T}}$ is paired with Narrow$^{\text{C}}$ ("equal": $\beta = 0.54$

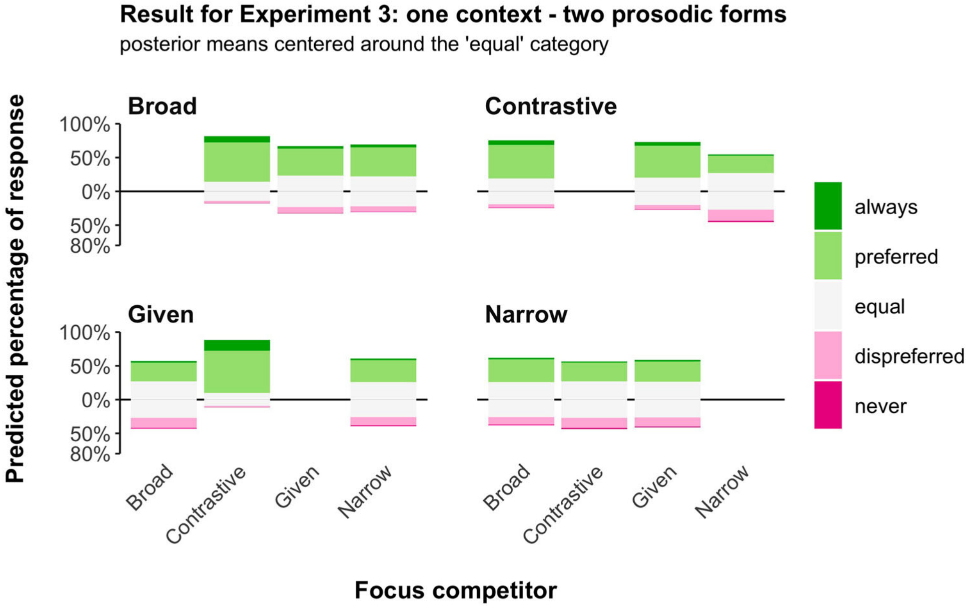

Figure 5. Stacked bar plots for the predicted probability of choosing one response over the others across target focus conditions and their accompanying focus competitors. Stacked bar plots are centred around the middle category ("equal") indicated by the solid horizontal line. Visual mass above the line indicates tendency to prefer the match between prosody and focus condition, mass below the line indicates tendency to not prefer the match.

[0.48,0.58]; "preferred": $\beta = 0.26$ [0.13,0.39]). Again, Broad and Contrastive focus categories elicit the strongest endorsements.

For Given^T, there is compelling evidence that Contrastive^C elicits fewer "equal" and more "preferred" and "always" responses ("equal": $\beta = 0.19$ [0.12,0.27]; "preferred": $\beta = 0.63$ [0.59,0.66]; "always": $\beta = 0.16$ [0.1,0.24]) than Broad^C ("equal": $\beta = 0.54$ [0.5,0.58]; "preferred": $\beta = 0.28$ [0.2,0.36]; "always": $\beta = 0.02$ [0.01,0.03]) and Narrow^C ("equal": $\beta = 0.52$ [0.46,0.57]; "preferred": $\beta = 0.32$ [0.22,0.42]; "always": $\beta = 0.16$ [0.1,0.24]), suggesting that listeners are most likely to endorse a Given^T prosody as correctly paired to its discourse context when the competitor pairing has Contrastive^C prosody.

Interestingly, Narrow^T did not elicit different responses across competitor categories. All three conditions seem to behave similarly and exhibit predominantly "equal" responses. This response pattern indicates that listeners endorse all prosodic patterns conditions as equally acceptable in pairings with the Narrow focus discourse context.

5.2. Results for experiment 3: Two contexts – one prosodic form

Figure 6 and Appendix 3 summarise the posterior distribution across conditions for the 2C-1P task. As opposed to the 1C-2P task, listeners do not show an overall tendency to endorse the stimuli pairings. There are generally stronger differences between focus conditions, with some eliciting responses predominantly on the negative end of the scale and others eliciting responses predominantly on the positive end of the scale. The generally weaker performance of listeners in this task compared to the 1C-2P task is in line with the results from Experiment 2.

For Broad^T, there is some evidence that Narrow^C elicits more "dispreferred" and "equal" and less "preferred" ratings ("dispreferred": $\beta = 0.46$ [0.39,0.53]; "equal": $\beta = 0.33$ [0.27,0.39]; "preferred": $\beta = 0.11$ [0.07,0.16]) than Contrastive^C ("dispreferred": $\beta = 0.24$ [0.16,0.32]; "equal": $\beta = 0.41$ [0.39,0.44]; "preferred": $\beta = 0.29$ [0.21,0.38]) and Given^C ("dispreferred": $\beta = 0.21$ [0.15,0.27]; "equal": $\beta = 0.41$ [0.38,0.43]; "preferred": $\beta = 0.32$ [0.24,0.4]). A general bias against Broad^T cannot be observed here (remember, in the forced choice 2C-1P task, Broad^T was systematically avoided as a possible response, whether congruent or incongruent on the trial).

The general bias in favour of the Narrow^C competitor remains apparent in experiment 3. When Narrow^C is available as a response option, listeners tend to prefer it over Broad^T. The bias towards Narrow responses can also be seen for Contrastive^T. When paired with Narrow^C, listeners selected more "dispreferred" and "equal" responses and less "preferred" responses ("dispreferred": $\beta = 0.36$ [0.26,0.46]; "equal": $\beta = 0.39$ [0.34,0.43]; "preferred": $\beta = 0.18$ [0.11,0.25]) than when Contrastive^T was paired with Given^C ("dispreferred": $\beta = 0.07$ [0.04,0.11]; "equal": $\beta = 0.25$ [0.17,0.33];

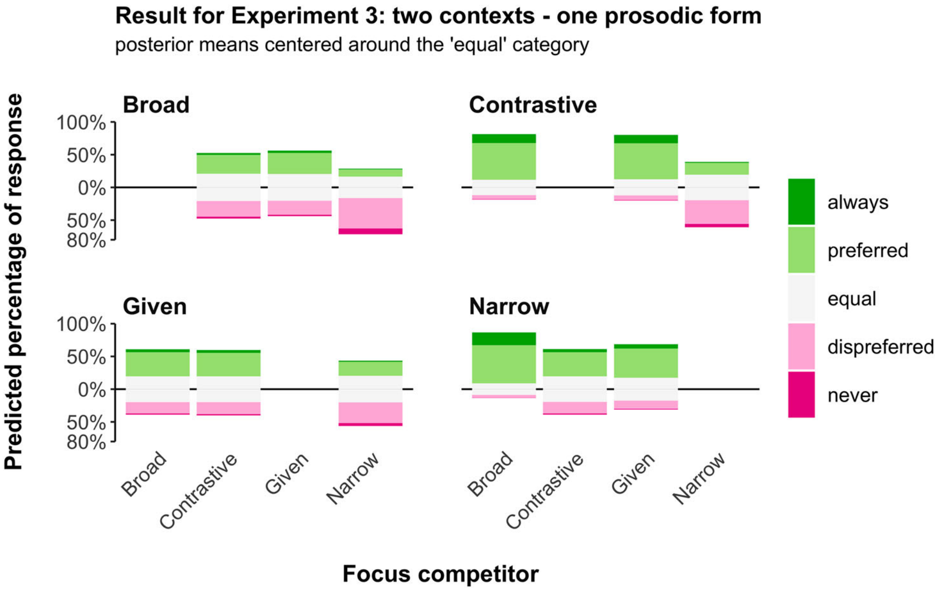

Figure 6. Stacked bar plots for the predicted probability of choosing one response over the others across target focus conditions and their accompanying focus competitors. Stacked bar plots are centred around the middle category ("equal") indicated by the solid horizontal line. Visual mass above the line indicates tendency to prefer the match between prosody and focus condition, mass below the line indicates the tendency to not prefer the match.

"preferred": $\beta = 0.55$ [0.49,0.61]) and Broad^{C} ("dispreferred": $\beta = 0.06$ [0.04,0.09]; "equal": $\beta = 0.23$ [0.17,0.3]; "preferred": $\beta = 0.56$ [0.51,0.6]).

Overall, Experiment 3 confirms the results from Experiments 1 and 2. Listeners can match different prosodic realisations to their speaker-intended focus categories, but listeners' performance differed across focus category pairs. In the 1C-2P task, endorsement was highest for the pairs {Contrastive, Given} as well as {Contrastive, Broad}. The acoustic overlap of their prosodic realisations explains some of these differences. The other pairs were not as well endorsed, including {Contrastive, Narrow} and {Given, Narrow}.

In the 2C-1P tasks, endorsement rates were generally more variable. In Experiment 3, where listeners are given more nuanced response options, the bias against matching a prosodic pattern to a Broad^{T} focus condition is not apparent anymore, but we do find evidence for the bias favouring matches to a Narrow focus condition, with weaker endorsement rates for Narrow^{C} and strong endorsement rates for Narrow^{T}.

In sum, the experiment with a 5-point response option qualitatively confirmed most of the results from Experiments 1 and 2. It also becomes clear that given more nuanced response dimensions, listeners turn out to be very liberal when it comes to acceptable matches between prosodic form and focus-related meaning established by discourse context.

6. General discussion

6.1. Summary

We have reported on three experiments to answer the question whether listeners perceive focus-related meaning on the basis of the prosodic form of an utterance. In Experiment 1, listeners had to decide which of two acoustic realisations matches a particular focus-related meaning established by the immediate discourse context (1C-2P). Listeners were able to distinguish different prosodic forms to match a certain focus category with above chance accuracy. Although listeners were able to match acoustic form and intended focus context, accuracy was rather low and performance varied strongly across different focus pairs. While some pairs of prosodically encoded focus categories seem to be more accurately distinguished (e.g. Contrastive vs. Broad, Contrastive vs. Given), other pairs elicited substantially worse performance, sometimes even failing to show above-chance accuracy (e.g. Given vs. Narrow).

In Experiment 2, listeners had to decide which of two focus categories specified by different discourse contexts is the best match to a particular acoustic prosodic form (2C-1P). This experiment uncovered interesting divergent results from Experiment 1, with listeners having greater difficulty matching question-answer pairs. We observed biases against selecting Broad focus as a match to any prosodic form, and favouring matches to Narrow focus to any prosodic form. These results suggest that listeners are influenced by other aspects of the stimuli than just the prosodic information in the signal (which they were clearly able to use in 1C-2P). As opposed to observed biases with Broad and Narrow focus prosody, listeners were able to assign Given and Contrastive prosodic realisations to their congruent discourse contexts.

Experiment 3 conceptually replicated Experiments 1 and 2 but using a 5-point scalar response option instead of a two-alternatives forced choice task. The results confirm what we have observed for the other experiments with one notable exception. The strong bias towards Broad focus vanishes in the 2C-1P task, suggesting that the bias towards Broad focus contexts only surfaces when listeners have to categorically decide for or against a context. When they have less restricted decision options, e.g. being able to choose that neither of the offered question-answer pairs is a better match, no bias against matches to the Broad focus context manifests anymore.

This is not true for the Narrow focus bias. Experiment 3 shows that listeners have a clear bias towards Narrow focus contexts, confirming that this focus type allows for a large variety of different prosodic realisations.

While we can confirm our hypothesis that listeners are sensitive to the acoustic prosodic expression of focus categories, there are two groups of questions that arise from our results: First, why are listeners' accuracies generally so low and why are some categories better distinguished than others? Second, why are listeners generally biased to match utterances with certain contexts but not with others?

6.2. Perceptual sensitivity is dependent on target and competitor category

The prosodic realisations of our stimuli are acoustically distinct (i.e. an LDA analysis can tease them apart with very high accuracy), so why do listeners have difficulties in mapping the speech signal onto speaker intentions?

One could argue that the low accuracy might be an artefact of the task, being artificial to some extent and devoid of (linguistic) functionality. Acoustic cues are more pronounced when the interlocutor is present (Breen et al., 2010; Buxó-Lugo, Toscano, & Watson, 2018; Turnbull, Royer, Ito, & Speer, 2017), when the speaker believes that the listener is distracted (Rosa, Finch, Bergeson, & Arnold, 2015), and when there is ambiguity in the context (Snedeker & Trueswell, 2003). Our model speaker produced her utterances in a context which is largely devoid of a communicative context, and listeners might not be able to access their entire knowledge about possible form-function mappings within the experiment. However, even in the experiment by Breen et al. (2010) which took great care in creating a functional communication situation between speakers and listeners, listeners still had difficulties mapping acoustic form onto intended focus type. This suggests that the low accuracy that we obtained is not necessarily an artefact of the task. Any explanation hinging on the artificial nature of the task also does not account for the fact that listeners can assign some prosodic forms to their intended focus context, in fact, with very high accuracy.

An alternative interpretation might be related to the amount of acoustic overlap between prosodic realisations of focus types. Concentrating on the f_0 maxima on the subject and the object constituent (as strongly related to the phonological pitch accent placement and pitch accent choice, see above), we can already see that some focus categories overlap more than others. For the f_0 maxima of the subject, the focus categories fall into two groups. Both the Broad and Contrastive groups, and the Given and Narrow groups overlap substantially. For the f_0 max of the object, Broad and Contrastive are actually well separated. Given and Narrow remain highly overlapping. These patterns reflect some of our 1C-2P results. Accuracy for Contrastive competing with Broad and Given was high, much higher than accuracy for Contrastive competing with Narrow. However, Given and Broad exhibit very well separated distributions in f_0 max, but elicit weaker accuracy, suggesting that there may be factors affecting their performance that go beyond simple acoustic overlap between categories.

Linguistic meaning is signalled by many temporally distributed cues throughout the discourse (e.g. Winter, 2014). Breen et al. (2010) showed that listeners' accuracies went up when the target sentences were preceded by the phrase "I heard that", suggesting that speakers prosodically signal focus categories on preceding syntactic material. Similarly, Xu and Xu (2005) found that focus categories are differentiated by both expanded pitch range on the focused constituents as well as post-focal compression on the lexical items following the focused constituent. Beyond distributed redundancy in the speech signal, non-verbal context might provide important disambiguating information. Speech communication does not happen in a void, but is accompanied by changes in body posture, head position, gaze, facial expressions, and manual gestures (e.g. Kendon, 2004; McNeill, 1992). For example, Krahmer and Swerts (2007) showed that Dutch speakers place more acoustic emphasis on words if their production is accompanied by a visual cue (eyebrow movement or head nod) and that subjects are more likely to perceive a word as prominent if accompanied by a visual cue. In an experiment such as the one that is the focus of the present analysis, in which speech stimuli are presented with very limited context, listeners have only a subset of information channels to make decisions about prosodically encoded meaning related to focus, leading to less certainty about their decisions. Some categories might benefit more or less from these contextual effects, accounting for category-specific performances.

Yet another aspect to consider is the inherent probabilistic nature of form-function mappings in prosody. One could argue that focus categories, as many other discourse functions, may not be discretely signalled by prosody in a deterministic way. In other words, listeners may be sensitive to prosodic cues, while recognising ambiguity in the mapping back to the speaker-intended meaning. Accumulating evidence reveals that intonation is characterised by a many-to-many-mapping between prosodic form and discourse function (Cangemi et al., 2015; Chodroff & Cole, 2018; Cruttenden, 1986; Grice et al., 2017; Peppé et al., 2000; Roettger, 2017; Turnbull, 2017). Specific prosodic forms are probabilistically associated with certain discourse functions. Language users have access to this knowledge which is reflected in (discretely) variable speech production patterns (one and the same speaker uses discretely different phonological forms to signal the same meaning) which results in observed flexibility in the comprehension of prosodically encoded discourse meaning in the lab (e.g. Roettger, 2017; Roettger & Grice, 2015).

In order to avoid making mappings that are different from those intended by the speaker, listeners need to adapt with respect to a given speaker (or a given context). The response data analysed here come from a series of experiments in which listeners rated as few as 18 utterances from a single speaker, offering only a slim basis for adaptation. A failure to adapt means that the listener's prior knowledge plays a greater role in speech perception. If listeners' prior beliefs of the form-function mapping for prosody is characterised by stochastic distributions rather than deterministic one-to-one relationships, that could account for some of the variability in the response patterns analysed here.

6.3. Listeners have biased expectations about suitable contexts

The rather low and inconsistent performance in mapping between prosodic form and discourse meaning might be a natural disposition of language users. The present study, like older studies on the perception of prosodic meaning, suggests that mapping an utterance onto a pragmatic meaning in the absence of a genuine communicative context is a difficult task and one that elicits highly variable performance from listeners. Nonetheless, and despite the inherent stochasticity of intonational form-function mappings, there are several studies showing that listeners rapidly integrate intonational information to anticipate speaker intentions (e.g. Dahan et al., 2002; Ito & Speer, 2008; Roettger & Stoeber, 2017; Watson et al., 2008; Weber et al., 2006). Listeners' ability to make use of bottom-up acoustic cues may be complemented by probabilistic knowledge about speaker production likelihoods, i.e. how likely the speaker is to use a particular prosodic form in order to express a particular discourse function (Buxo-Lugo, 2017; Buxó-Lugo & Watson, 2016; Kurumada et al., 2014; Roettger & Franke, 2018a, 2018b).

For example, in Roettger and Franke (2018a, 2018b), listeners were exposed to two intonation contours. These contours exhibited early intonational cues to speaker intentions, i.e. cues that become available before the lexical content disambiguates between competing interpretations of discourse meaning. Roettger and Franke showed that the assumed production likelihood of a prosodic cue predicted listeners' anticipatory behaviour at the beginning of the experiment as well as its development through exposure to confirming or disconfirming observations. In other words, when exposed to stochastically confirming or disconfirming form-function mappings, listeners adapt to what extent they predictively use an intonational cue. If listeners learn that an intonational cue (e.g. a particular pitch accent) is uninformative, they appear to weigh the informational value of that cue less heavily (see also Kurumada et al., 2014).[4] Roettger and Franke's results are in line with the assumption that language users have probabilistic knowledge about the stochastic co-occurrence of prosodic form and discourse function.

Coming back to the present findings, the above insights may offer an explanation as to why listeners are biased to (erroneously) reject broad focus and to (erroneously) accept narrow focus in the 2C-1P task. The broad focus question was a question like "What has happened?". This (or similar questions) are often used to elicit broad focus in the experimental literature. Semantically, this question does not pre-activate any discourse relations and allows for an out-of-the-blue interpretation. However, this discourse context is pragmatically very rare. We rarely encounter out-of-the-blue scenarios without any prior knowledge about the discourse, thus the likelihood of a speaker expressing (truly) broad focus is arguably very low. As opposed to that, the given context, i.e. repeating the previously heard proposition, and contrastive focus, i.e. correcting the previously heard proposition, are very common discourse scenarios, albeit occurring in very specific discourse contexts. Finally, narrowly focusing a constituent is arguably a very general pragmatic function that applies to many different discourse contexts. We encounter a narrow focus context very often, thus the likelihood of a speaker expressing narrow focus is arguably very high. For exposition purposes, let us assume that narrow and broad focus are not prosodically differentiated (so any intonational cue (I) has the same probability (P) of expressing Narrow (N) and Broad (B) focus, i.e. $P(I|B) = P(I|N)$). If the prior belief about the likelihood of a discourse function is asymmetric, i.e. $P(B) < P(N)$, listeners would believe that narrow focus is more likely, i.e. the probability of a narrow focus interpretation given any intonational cues would be higher than the probability of a broad focus interpretation given the same intonational cues. This relationship can be expressed via Bayes Rule, cf. (1):

$$\frac{P(N|I)}{P(B|I)} = \frac{P(I|N)}{P(I|B)}\frac{P(N)}{P(B)} > \frac{P(I|B)}{P(I|N)}\frac{P(B)}{P(N)} = \frac{P(B|I)}{P(N|I)} \qquad (1)$$

This proposal is in line with a rational analysis approach (Anderson, 1990) to speech perception (e.g. Clayards et al., 2008; Kleinschmidt & Jaeger, 2015), assuming that prosodic perception and processing can be conceptualised as a process of *inference under uncertainty*: listeners know that certain discourse functions are realised as a distribution of acoustic cues and the listener probabilistically infers how likely any given speaker intention is, taking into account both their knowledge about stochastic cue distributions as well as their knowledge about speaker and context. We want to emphasise that this is an ad-hoc explanation that remains speculative until further investigations. We believe, however, that this explanation offers an insightful perspective on previous findings in general and our findings in particular.

7. Conclusion

The prosodic modulation of speech is a tremendously important aspect of human language. However, our knowledge as to how language users interpret prosody to guide intention recognition is still surprisingly small.

The present paper contributes to this knowledge. We have presented evidence that listeners can use prosodic information to infer the intended information structure of an utterance, even in a laboratory setting that is devoid of contextual information. These results complement the existing literature on American English in that they clearly show listener's ability to discriminate prosodic forms intended by the speaker to signal focus types (e.g. Breen et al., 2010; Gussenhoven, 1983; Welby, 2003). Our study further contributes to research on prosody and meaning in general in that our 2C-1P tasks allow us to uncover certain meaning biases in how listeners associate prosodic forms with focus-related discourse meaning. The experimental tasks used here may tap into comprehension processes that are not only informed by acoustic information but also by listeners' prior knowledge of the contextual probability of a prosodic form.

More clearly than in previous studies, the experiments presented in this study suggest a high degree of overlap in the pairing of prosodic form and information structure categories, with some prosodically encoded focus types being more accurately associated with discourse contexts than others. These differences may be related to different degrees of acoustic / perceptual overlap between the prosodic categories. Although we did not investigate a representative sample of production data, we have discussed idiosyncratic patterns of our model speaker for whom some categories may be more or less overlapping with regard to relevant phonetic dimensions.

In addition to acoustic prosodic properties of the intended focus types, our data suggests additional factors contributing to listeners' mappings of form onto function: Listeners appear to be influenced by their probabilistic knowledge about how likely a speaker is to produce a certain prosodic form and how likely this form will be used as intended by a speaker to communicate a certain discourse function. The latter explanation can account for the observed meaning biases and is in line with recent studies on intonational processing (Buxó-Lugo & Watson, 2016; Kurumada et al., 2014; Roettger & Franke, 2018a, 2018b) and speech perception in general (e.g. Clayards et al., 2008; Kleinschmidt & Jaeger, 2015; Kleinschmidt et al., 2018; Norris et al., 2003). This explanation, although grounded in recent experimental studies, remains speculative and should merely serve as a departure point for future studies working on the mapping of prosody and meaning.

We conclude that listeners infer speaker intentions based on both bottom-up acoustic cues and top-down probabilistic expectations about likely discourse contexts.

Notes

1. The findings from these experiments are also discussed in Mahrt (2018), with qualitative comparisons across experimental conditions.
2. The ToBI labels represent the one or two most frequent pitch accents produced on the subject and object nouns, over the nine sentence stimuli in each focus condition.
3. Our decision to administer both tasks (1C-2P, 2C-1P) to the same participants in Experiment 3, rather than run two separate experiments with the expanded, scalar response set, were driven by practical constraints of time and money. We were also interested in testing the feasibility of combining both tasks in one experiment, to approximate the design of earlier experiments, testing categorical perception of phoneme contrasts, in which identification and discrimination experiments are administered to the same participants. Mahrt (2018, p. 30) reports on a pilot experiment using both tasks, involving 45 subjects recruited from Mechanical Turk. In the pilot experiment, a third of the participants first did the identification task followed by the discrimination task, another third did the tasks in the opposite order, and the final third did them with the tasks interleaved in random sequence over trials. The same set of stimuli were presented to all participants. The results of the pilot showed that the task done second resulted in higher accuracy than when it was performed first, with intermediate accuracy for participants in the interleaved tasks condition. On the basis of these findings, Experiment 3 adopted the interleaved task design.
4. This adaptive behaviour is in line with language users adapting readily to their immediate local context in syntax (e.g. Fine, Jaeger, Farmer, & Qian, 2013; Jaeger & Snider, 2013), pragmatics (e.g. Grodner & Sedivy, 2011; Yildirim, Degen, Tanenhaus, & Jaeger, 2016), and, most importantly, in speech (e.g. Kleinschmidt & Jaeger, 2015; Norris et al., 2003).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Timo B. Roettger http://orcid.org/0000-0003-1400-2739

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Earlbaum.

Audacity Team. (2015). *Audacity (r): Free audio editor and recorder* [computer program]. Version 2.1.0.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv* Preprint *ArXiv*:1506.04967.

Baumann, S. (2006). *The intonation of givenness - Evidence from German*. PhD thesis, Saarland University. Linguistische Arbeiten 508. Tübingen: Niemeyer.

Baumann, S., Röhr, C. T., & Grice, M. (2015). Prosodische (De-)Kodierung des Informationsstatus im Deutschen. *Zeitschrift für Sprachwissenschaft*, *34*(1), 1–42.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(1), 255–309.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098.

Brown, G. (1983). Intonation, the categories given/new and other sorts of knowledge. In A. Cutler & D. R. Ladd (Eds.), *Prosodic function and prosodic representation* (pp. 67–78). Cambridge: Cambridge University Press.

Büring, D. (2006). Intonation und Informationsstruktur. In H. Blühdorn, E. Breindle, & U. H. Waßner (Eds.), *Text - Verstehen. Grammatik und darüber hinaus* (pp. 144–163). Berlin: de Gruyter.

Büring, D. (2012). Focus and intonation. In G. Russell & D. Graff Fara (Eds.), *The Routledge companion to the philosophy of language* (pp. 103–115). London: Routledge.

Bürkner, P.-C. (2016). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Buxo-Lugo, A. F. (2017). *Communicative context, expectations, and adaptation in prosodic production and comprehension* (Doctoral dissertation). University of Illinois at Urbana-Champaign.

Buxó-Lugo, A. F., Toscano, J. C., & Watson, D. G. (2018). Effects of participant engagement on prosodic prominence. *Discourse Processes*, *55*(3), 305–323.

Buxó-Lugo, A. F., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, *90*, 1–13.

Calhoun, S. (2006). *Intonation and information structure in English* (Doctoral dissertation, PhD thesis). University of Edinburgh.

Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, *40*(2), 329–349.

Cangemi, F., & Grice, M. (2016). The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Laboratory Phonology*, *7*(1), 1–20.

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual differences in speech production and perception* (pp. 123–145). Frankfurt a. M.: Peter Lang.

Chafe, W. (1987). Cognitive constraints on information flow. *Coherence and Grounding in Discourse*, *11*, 21–51.

Chodroff, E., & Cole, J. (2018). Information structure, affect, and prenuclear prominence in American English. *Proceedings of Interspeech 2018*, 1848–1852.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

Cooper, W., Eady, S., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142–2156.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One*, *8*(10), e77661.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768.

Grice, M., Ridouane, R., & Roettger, T. B. (2015). Tonal association in Tashlhiyt Berber: Evidence from polar questions and contrastive statements. *Phonology*, *32*(2), 241–266.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107.

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (pp. 239–272). Cambridge, MA: MIT Press.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech*, *26*(1), 61–80.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83.

Jaynes, E. T., & Kempthorne, O. (1976). Confidence intervals vs. Bayesian intervals. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science* (Vol. 6b, pp. 175–257). Dordrecht: Springer.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kimball, A. E., Shantz, K., Eager, C., & Roy, J. (2018). Beyond maximal random effects for logistic regression: Moving past convergence errors. *Journal of Quantitative Linguistics*, 1–25.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Kleinschmidt, D. F., Weatherholtz, K., & Florian Jaeger, T. (2018). Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science*, *10*, 818–834.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57* (3), 396–414.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. (2014). Rapid adaptation in online pragmatic interpretation of contrastive prosody. In *Proceedings of the annual meetinf of the cognitive science society, 36*. Austin, TX: Cognitive Science Society.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*(1), 81–112.

Mahrt, T. (2016). *LMEDS: Language markup and experimental design software*. Retrieved from https://github.com/timmahrt/LMEDS

Mahrt, T. (2018). *Acoustic cues for the perception of the information status of words in speech*. Urbana, IL: University of Illinois at Urbana-Champaign.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago press.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Peppé, S., Maxim, J., & Wells, B. (2000). Prosodic variation in southern British English. *Language and Speech*, *43*(3), 309–334.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Bloomington, IN: MIT.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Roettger, T. B. (2017). *Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments*. Berlin: Language Science Press.

Roettger, T. B., & Franke, M. (2018a). Dynamic speech adaptation to unreliable cues during intonational processing. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Proceedings of annual meeting of the cognitive science society, 40* (pp. 966–971). Austin, TX: Cognitive Science Society.

Roettger, T. B., & Franke, M. (2018b). *Evidential strength of intonational cues and rational adaptation to (un-)reliable intonation*. Preprint at PsyArXiv: Retrieved from https://psyarxiv.com/awp87

Roettger, T. B., & Grice, M. (2015). The role of high pitch in Tashlhiyt Tamazight (Berber): Evidence from production and perception. *Journal of Phonetics*, *51*(1), 36–49.

Roettger, T. B., & Stoeber, M. (2017). Manual response dynamics reflect rapid integration of intonational information during reference resolution. In G. Gunzelmannn, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of annual meeting of the cognitive science society, 39* (pp. 3010–3015). Austin, TX: Cognitive Science Society.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116.

Rosa, E. C., Finch, K. H., Bergeson, M., & Arnold, J. E. (2015). The effects of addressee attention on prosodic prominence. *Language, Cognition and Neuroscience*, *30*(1–2), 48–56.

Rump, H. H., & Collier, R. (1996). Focus conditions and the prominence of pitch-accented syllables. *Language and Speech*, *39*(1), 1–17.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Cambridge, MA: Blackwell.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130.

Turnbull, R. (2017). The role of predictability in intonational variability. *Language and Speech*, *60*(1), 123–153.

Turnbull, R., Royer, A. J., Ito, K., & Speer, S. R. (2017). Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience*, *32*(8), 1017–1033.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S (Fourth)*. New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H. *Cognitive Science*, *32*(7), 1232–1244.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392.

Welby, P. (2003). Effects of pitch accent position, type, and status on focus projection. *Language and Speech*, *46*(1), 53–81.

Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays*, *36*(10), 960–967.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.

Yildirim, I., Degen, J., Tanenhaus, M. K., & Jaeger, T. F. (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language*, *87*, 128–143.

Yoon, T.-J. (2010). Speaker consistency in the realization of prosodic prominence in the Boston University Radio Speech Corpus. In *Proceeding of 5th speech prosody, Chicago*.