

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Manual Response Dynamics Reflect Rapid Integration of Intonational Information during Reference Resolution

Permalink

<https://escholarship.org/uc/item/4rc3570t>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors

Roettger, Timo B.

Stoeber, Mathias

Publication Date

2017

Peer reviewed

Manual Response Dynamics Reflect Rapid Integration of Intonational Information during Reference Resolution

Timo B. Roettger & Mathias Stoeber

timo.roettger@uni-koeln.de, m.stoeber@uni-koeln.de

Department of Linguistics – Phonetics, University of Cologne, Herbert-Lewin-Straße 6
D-50931 Cologne, Germany

Abstract

Intonation plays an integral role in comprehending spoken language. It encodes post-lexical pragmatic functions such as sentence modality and discourse contexts. The present experiment investigates how and when listeners integrate intonational information to anticipate reference resolution. While most work on the real-time processes of intonation-based intention recognition has utilized eye tracking, the present study uses the mouse tracking paradigm, a valuable complementary method to investigate the time course of speech processing. Participants had to choose an interpretation based on pre-recorded instructions containing different intonation contours. Recordings of the x,y coordinates of participants' computer mouse movements reveal that listeners integrate intonational information rapidly as soon as they become available and anticipate potential referential interpretations early on.

Keywords: intonation, reference resolution, mouse tracking

Introduction

During the perception of an unfolding speech signal, listeners use acoustic information to guide their interpretation of what a speaker intends to communicate. This process can take place long before disambiguating lexical information becomes available, allowing the listener to make rapid inferences about what a speaker intends to say, even if these inferences are based on partial information.

Intonation plays an integral role in this interpretation process. Among other things, intonation is commonly used to express discourse relations such as givenness and contrastiveness (e.g. Ladd 2008). Intonational acoustic events such as pitch accents have been shown to consistently encode the discourse status of referents. For example, in German or English, a high rising pitch accent generally signals new information, while for example deaccentuation (i.e. the absence of a pitch accent) tends to signal given information (e.g. Fery & Kügler 2008, Cangemi *et al.* 2015 *inter alia*).

While much work has been done on how intonational events encode discourse relations, there is only little work on how and when listeners integrate this acoustic information with the relevant discourse information.

To fully understand intonation-based intention recognition, it is necessary to use experimental techniques that measure the real-time integration of intonational information to resolve temporally ambiguous interpretations. While eye tracking experiments have advanced our knowledge about these processes tremendously (e.g. Dahan *et al.* 2002, Weber *et al.* 2006, Watson *et al.* 2006, Kurumada *et al.* 2014), it has been pointed out that the nature of oculomotor patterns constitutes a limitation of the eye-tracking paradigm (e.g. Spivey *et al.*

2005, Dale *et al.* 2007). Eye-movement data is characterised by ballistic “jumps” of the eye. Only by averaging over many trials can a pseudo-continuous trajectory be calculated, which is then interpretable as evidence for a continuous comprehension process.

This potential methodological shortcoming can be overcome by measuring another form of movement behaviour: the movements of hands. Over the last decade, it has been demonstrated that continuous nonlinear trajectories recorded from the streaming of x,y coordinates of computer mouse movements can serve as an informative indicator of cognitive processes (e.g. Spivey *et al.* 2005, Magnuson 2005). Even though mouse tracking has been applied to diverse phenomena in cognitive science, its usefulness for speech processing research has been somewhat neglected. This paper provides evidence that mouse tracking is suitable to unravel real-time dynamics of speech processing beyond lexical and phonemic processing (see also Tomlinson & Bott 2013 and Warren 2017).

Real Time Integration of Intonation

Several studies have demonstrated that comprehenders can rapidly integrate intonational information to map an utterance containing referential expressions onto intended referents. These studies have focused on the discourse status of referents, i.e. whether an item has or has not already been mentioned or is explicitly contrasted to another referent.

Dahan *et al.* (2002) utilised the visual world paradigm in which specific items and geometrical shapes were distributed in a grid. Upon hearing specific auditory instructions listeners had to move the objects above or below the shapes. In one of their experiments, subjects heard a trigger sentence such as “Put the candle below the triangle.” The object “candle” and the location “below the triangle” were thus introduced to the listener as given information. After the trigger sentence, listeners heard the critical instruction, either referring to the given object (“candle”) or a lexical competitor which shares its word onset with the given object (here “candy”). When the target word was deaccented, listeners' eye movements revealed significantly more fixations to the already given object before the lexical disambiguation was available. Conversely, when the target word was accented, there were more fixations to the competitor.

Weber *et al.* (2006) extended these findings by showing that the presence or absence of a contrastive pitch accent on a modifying adjective allows listeners' anticipation of contrastivity of the noun (see also Watson *et al.* 2006).

Kurumada *et al.* (2014) examined the time course of the construction “It looks like an X” pronounced with either a high pitch accent on the final noun followed by a low boundary tone (e.g. it looks like a ZEBRA), or a contrastive high rising pitch accent on the verb and a rising boundary tone, a contour that can support contrastive inference, (e.g. it LOOKS like a zebra (but it is not)). They found that listeners integrate the pitch accent information on the verb to anticipate the status of the target referent.

These reported effects are consistent with the hypothesis that listeners integrate intonational information rapidly as soon as it becomes available and anticipate potential referential interpretations early on.

The contributions of the present paper to this literature are twofold: On the one hand, we aim to replicate previous findings showing that listeners take up intonational information rapidly to anticipate pragmatic interpretations using mouse tracking. On the other hand, we aim to prove that continuous response tracking can provide valuable insights into the real-time comprehension of utterance-long speech signals.

The Present Study

The present study investigates intonation-based intention recognition in German using the mouse tracking paradigm. It is hypothesised that listeners integrate intonational information to anticipate referential ambiguity early on and that this anticipation is reflected in the dynamics of their mouse trajectories during response selection. In line with standards of reproducible research, all materials (including audio and visual stimuli), scripts, and raw data are available here: <https://osf.io/n79x3>

Methodology

Participants had to choose visually presented response alternatives corresponding to pre-recorded speech files in a two-alternative forced choice design. Stimuli differed with respect to the available discourse context and the intonationally encoded information status of referents, enabling us to investigate the real-time integration of intonational information during reference resolution.

Experimental Set-up Participants were seated in front of a MacBook Pro 3.1 GHz Intel Core i7 with a display resolution of 1280x800. They controlled the experiment via a Logitech B100 corded USB Mouse. Cursor acceleration was made linear and cursor speed was slowed down using the CursorSense© application (version 1.32).

Participants and Procedure Ten native speakers of German (five male, five female) with an average age of 30.3 years (SD = 4.9) participated in this experiment. All of them had normal or corrected-to-normal vision.

Participants were told about two different fantasy creatures which were introduced as ‘wuggies’. These wuggies were displayed as having picked up certain real world objects such as a pear or a violin. The two wuggies differed in colour

(“blau” ‘blue’ vs. “gelb” ‘yellow’) and there were 10 different objects that the wuggies could pick up (bee, chicken, fork, marble, pants, pear, rose, saw, vase, violin).

On each trial, participants were exposed to a question screen that either did or did not provide a specific discourse context, followed by a response screen in which participants had to choose visually presented response alternatives depending on an auditorily presented sentence.

On the question screen, participants either heard nothing or they heard a question such as (1):

- (1) “Hat der gelbe Wuggy die Geige aufgesammelt?”
‘Did the yellow wuggy pick up the violin?’

The question provided a discourse context with certain elements being activated as given for the participant (here: the yellow wuggy and the violin). The question screen was visible for 2500 ms.

Following the question screen, participants saw two visually presented response alternatives, each depicting a wuggy carrying an object. After 1000 ms, a yellow circle appeared at the bottom centre of the screen. Participants were instructed to click on the yellow circle to initiate playback of an audio recording. The audio recording was a statement specifying which wuggy has picked up which object, e.g., “Der gelbe Wuggy hat die Geige aufgesammelt.” ‘The yellow wuggy has picked up the violin.’ Participants were instructed to move their mouse immediately upwards after clicking the initiation button and choose the respective response alternative as soon as they could.

After each response selection, the screen was left blank for a 1000 ms inter-stimulus interval.

Prior to the beginning of the experimental trials, participants were given 36 practice trials to familiarise themselves with the paradigm.

Speech Material There were two sets of acoustic stimuli: questions providing a discourse context presented on the question screen and statements triggering participants’ responses on the response screen. There were twenty different questions for all possible combinations of wuggies and objects (two wuggies × ten objects).

Likewise, there were twenty different statements, which were produced with four different intonation contours. Based on the question, i.e. the discourse context, and the visual scene at hand, statements differed with regard to the information status of the relevant constituents of the sentence: The question in (1) (“Hat der gelbe Wuggy die Geige aufgesammelt?”) asks for confirmation that the proposition (including the identity of the subject and object) is true. Now consider the following answers (2-4):

- (2) Der gelbe Wuggy hat die Geige aufgesammelt.
‘The yellow wuggy has collected the violin.’
- (3) Der gelbe Wuggy hat die Birne aufgesammelt.
‘The yellow wuggy has collected the pear.’
- (4) Der blaue Wuggy hat die Geige aufgesammelt.
‘The blue wuggy has collected the violin.’

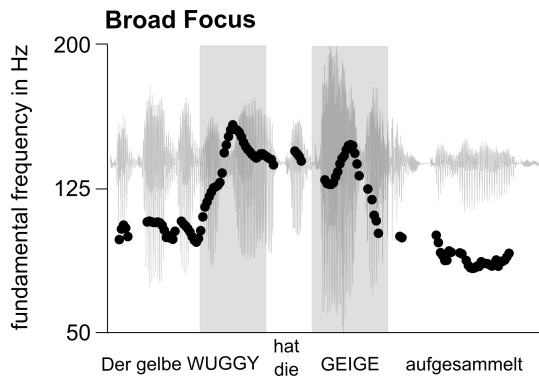


Figure 1: Representative waveform and f0 contour for a statement produced with a rising accent on “Wuggy” and a falling accent on “Geige”, a typical contour for broad focus. Accented words are highlighted with grey boxes.

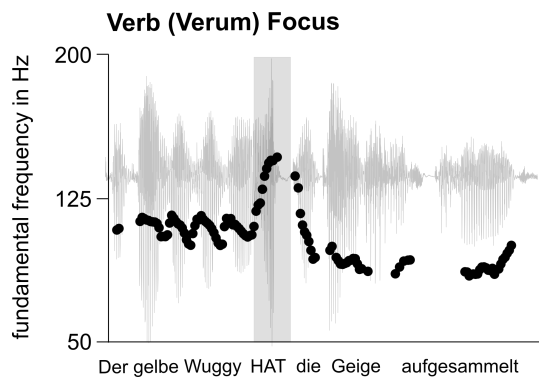


Figure 2: Representative waveform and f0 contour for a statement produced with a rising accent on the auxiliary “hat”, typically used to indicate verum focus. The accented word is highlighted with a grey box.

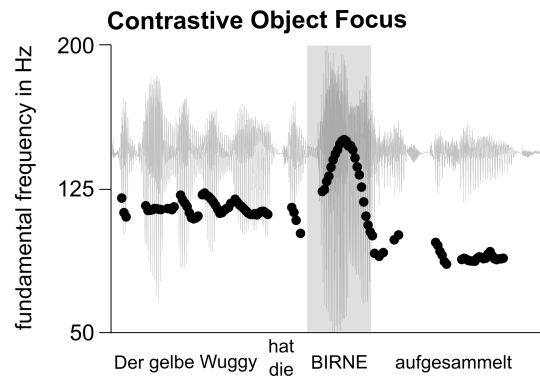


Figure 3: Representative waveform and f0 contour for a statement produced with a rising accent on the referent “Birne”, typically used to indicate contrastive focus. The accented word is highlighted with a grey box.

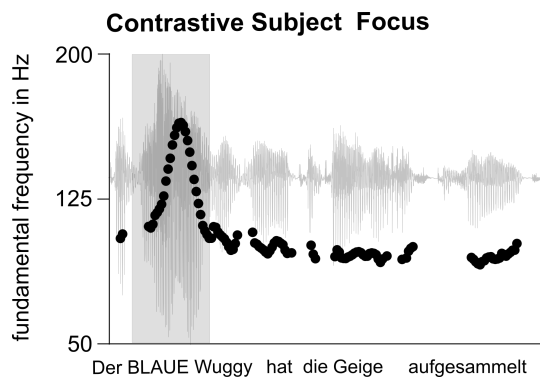


Figure 4: Representative waveform and f0 contour for a statement produced with a rising accent on the subject modifier “blaue”, typically used to indicate contrastive focus. The accented word is highlighted with a grey box.

Dependent on the discourse context (here: whether there is question or not, and the question being asked), the answers in (2-4) are realised with different intonation contours (Fery & Kügler 2008, Cangemi *et al.* 2015). If there is no discourse context available, both the subject and the object are new information in (2) (often referred to as broad focus) which can be prosodically encoded by specific pitch accents on both constituents. A common contour in these cases is a rising accent on the subject, followed by a high stretch of f0 and a high or falling accent on the object (cf. Figure 1).

Alternatively, if there is a relevant discourse context such as the question in (1), the utterance in (2) can prosodically emphasise that the proposition of the question is true. This can be indicated, for example, by verum focus, which manifests itself here in the form of a high rising accent on the auxiliary (“hat”, cf. Figure 2).

In contrast, the answers in (3) and (4) correct the proposition of the question. In (3), “Birne” is explicitly contrasted with “Geige”, typically expressed by an intonation contour with a high rising accent on “Birne” (cf. Figure 3). In (4), “blaue Wuggy” contrasts with “gelbe Wuggy” in the question. In this context, a high rising accent on “blaue” and no

accent on “Geige” is typically found (cf. Figure 4).

All acoustic stimuli were produced by a trained phonetician in a sound-attenuated booth at the Institute of Phonetics in Cologne with a headset microphone (AKG C420) using 48 kHz/16 bit sampling. The average stimulus duration of the trigger sentences was 1993 ms.

Visual Stimuli The pictures of the fantasy creatures were taken from a hand drawn set developed and used by van de Vijver & Baer-Henney (2014). The pictures of objects were taken from the BOSS corpus (Brodeur *et al.* 2010).

Response alternatives of critical trials differed visually by the identity of the referent only (e.g. yellow wuggy carrying a pear vs. yellow wuggy carrying a violin). In addition to the critical trials, we included the same number of filler trials, in which response alternatives differed visually by the colour of the wuggy only (e.g. yellow wuggy carrying a pear vs. blue wuggy carrying a pear), or by both the colour of the wuggy and the identity of the object (e.g. yellow wuggy carrying a pear vs. blue wuggy carrying a violin). These visual contrasts were introduced to ensure that participants do not simply

learn to anticipate certain combinations of questions and visual contrasts, disregarding the acoustic information.

Stimuli Presentation and Predictions There were four different experimental conditions: In the broad focus condition, participants did not receive a question and had to respond to a broad focus statement (cf. Figure 1). Since participants had no discourse context available, they had to rely on lexical information only. It is expected that the mouse movements during reference resolution do not change until the lexical information becomes available (the onset of “Geige” in example 2).

In the other three conditions, participants received a question and were thus able to integrate the given discourse context with the intonational information encoding the information status of the referents. Participants saw a already mentioned, given object (here: “Geige”) and a new object (here: “Birne”).

In the object focus condition, the pitch accent on the object indicates the contrastive nature of the object. The available pitch accent information becomes available simultaneously with the lexical information, i.e. the rise in pitch starts at the onset of the word (cf. Figure 3). Assuming that the pitch accent information primarily cues contrastivity, we do not expect listeners to anticipate the referent, i.e. the broad and object condition should not differ.

In the verb focus condition, the pitch accent on the verb indicates verum focus, i.e. signalling that the proposition of the question is true implying that the statement contains the already mentioned object (here: “Geige”). As soon as the intonational information on “hat” becomes available, listeners are expected to integrate this information, enabling reference resolution before the lexical information becomes available.

In the subject focus condition, the pitch accent on the subject modifier indicates the contrastive nature of the subject. This information enables an early inference towards the given nature of the object which only occurs later in the utterance, making reference resolution possible very early on.

Left/right placement of target vs. distractor response alternatives was counterbalanced within participants.

Analysis The x, y screen coordinates of the computer mouse were sampled at 100 Hz using the mousetrap plugin (Kieslich & Henninger 2016) implemented in the open source experimental software OpenSesame (Mathôt *et al.* 2012). Trajectories were processed with the package *mousetrap* (Kieslich *et al.* 2017) using the statistical software R (2016).

There was a total of 80 target trials, for a grand total of 800 trajectories across participants (200 per condition). Overall, 4.36 % of trials with incorrect responses and 0.45 % of trials with initiation times greater than 500 ms were discarded. Additionally, 1.67 % of trials were excluded due to movement behaviour that violated instructions (loops, reaching the top of the screen before response selection).

For each of the remaining trials, we computed two measurements based on time- and space-normalised trajectories: First, we collected overall reaction times (RT) measured from

the initiation click up until reaching the target response. This serves as a latency baseline. Second, we measured the area under the curve (AUC) operationalised by the geometric area between the observed trajectory and an idealised straight-line trajectory drawn from the start and end points (Freeman & Ambady 2010). A greater AUC is indicative of greater response competition between target and competitor during response selection.

We analysed data using hierarchical linear models using R and the package *lme4* (Bates *et al.* 2015), *afex* (Singmann *et al.* 2016), and *lmerTest* (Kuznetsova, Brockhoff, & Christensen 2016). Discourse condition (broad, object, verb, subject) was included as a fixed effect. Participants were specified as by-condition random slopes and referents were specified as random intercepts.

Results and Discussion

Inspection of time- and space-normalised horizontal trajectories over time (cf. Figure 5) suggests that trajectories were characterised by initially gravitating toward the midpoint between response alternatives (horizontal cursor position = 0) before eventually curving towards the target response (horizontal cursor position = -1).

Focus conditions elicited similarly-shaped trajectories that mainly differed with respect to their temporal characteristics. Not surprisingly, conditions differed in their overall response latency, measured from clicking the initiation circle to reaching the target response area ($\chi^2(3)=19.6$, $p=0.0002$) with the broad condition being the overall slowest ($\beta=1578$ ms, $SE=40.7$) followed by the object condition ($\beta=1457$ ms, $SE=49.5$), the verb condition ($\beta=1367$ ms, $SE=51.9$), and the subject condition ($\beta=1121$ ms, $SE=93.8$) (cf. Figure 6, Table 1). Pairwise comparisons reveal significant differences between all four conditions. The earlier the relevant intonational cue in the acoustic signal, the faster listeners selected a response. Moreover, the difference between broad condition and object condition suggests that the integration of discourse context and intonation facilitated reference resolution (object) in contrast to cases without available discourse context (broad).

These overall response latencies were neatly reflected in early moments of direction change: The broad condition started curving towards the target response around 300 ms after lexical disambiguation (dashed line in Figure 5). This time-lag can be interpreted as the time it takes for listeners' movements to be affected by the relevant acoustic cue of the lexical item.

The subject condition elicited response trajectories that deviated towards the target response very early in the signal, after having heard the contrastive pitch accent on the subject modifier. As opposed to that, the verb condition started curving towards the target shortly after the acoustic onset of the referent suggesting that integrating intonational information of the verum focus led to an immediate anticipation of the target referent before lexical disambiguation has taken place.

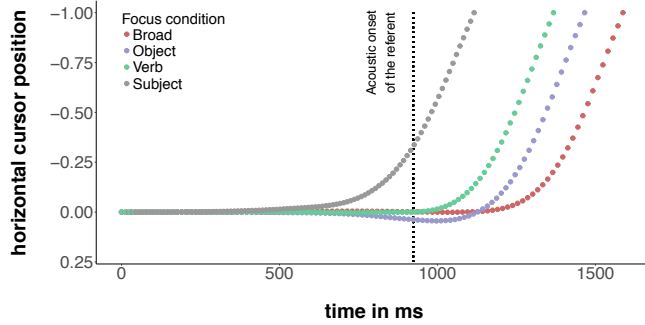


Figure 5: Horizontal cursor position of mouse trajectories plotted over time for broad, object, verb, and subject condition. Dashed line indicates the averaged acoustic onset of the critical lexical item.

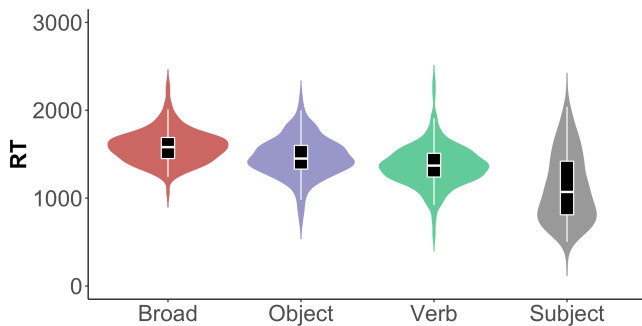


Figure 6: Violin plots of overall response latency (RT) of response selection.

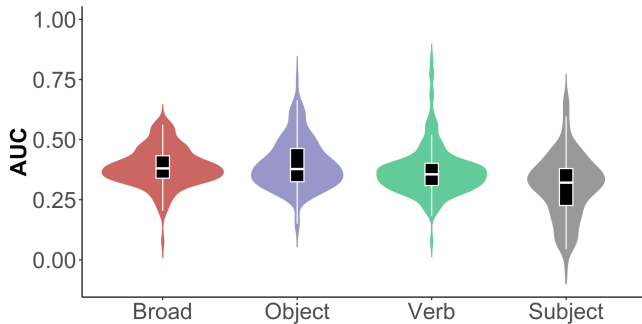


Figure 7: Violin plots of area under the curve (AUC) values of response selection.

Crucially, the object condition started its curvature towards the target less than 200 ms after the point of lexical disambiguation. Taking the time lag of the broad condition into account, it becomes clear that even the object condition elicited trajectories that started curving towards the target response before lexical disambiguation had taken place. Listeners’ anticipation in the object conditions seems puzzling. Both the lexical cue (onset of disambiguating phones) and the intonational cue (onset of the rising pitch movement) become available in the signal at the same time, i.e. the onset of the referential expression. The question arises as to how listeners anticipate the referent in the object condition. We propose two

possible answers to this question: On the one hand, linguistic functions are expressed by multiple acoustic cues distributed throughout the signal. Listeners might have picked up acoustic evidence indicating the contrastive nature of the referent *before* the pitch accent information had become available. On the other hand, within the microcosms of the experiment, listeners might have been able to anticipate the referent based on the *absence* of contradicting information. In other words, listeners did not hear a pitch accent on either the subject modifier nor the auxiliary, leading them to the conclusion that the object must be contrastive.

Overall, the observed patterns suggest that the integration of intonational information and discourse context facilitated reference resolution due to successful anticipation.

Table 1: Descriptive and inferential summary statistics for RT and AUC for each focus condition.

Condition	RT			AUC		
	mean	est	SE	mean	est	SE
Broad	1587	1578	40.7	0.39	0.39	0.02
Object	1466	1457	49.5	0.4	0.4	0.02
Verb	1367	1363	51.9	0.37	0.37	0.01
Subject	1116	1121	93.8	0.31	0.31	0.03

Beyond these temporal operationalisations, results for area under the curve measurements (AUC) indicated that conditions differed in overall attraction of trajectories towards the competitor ($\chi^2(3)=12.4$, $p=0.006$) with the object condition exhibiting the greatest AUC ($\beta=0.4$, $SE=0.02$), followed by the broad condition ($\beta=0.39$, $SE=0.02$), the verb condition ($\beta=0.37$, $SE=0.02$), and the subject condition ($\beta=0.31$, $SE=0.03$) (cf. Figure 7, Table 1). Not surprisingly, the earlier the relevant intonational cue in the acoustic signal, the less curvature towards the competitor was found. Importantly, albeit highly correlated, AUC and RT were not a direct mirror image of each other. While the broad condition is clearly the slowest, it was not the condition with the greatest AUC indicating that these measures reflect two different aspects of the response selection process: Overall latency and response competition.

General discussion

The present study investigated intonation-based intention recognition in German using the mouse tracking paradigm. Listeners were exposed to different discourse contexts and different intonational patterns encoding the discourse status of referents. Analyses of continuous computer mouse movements during response selection suggest that listeners integrated intonational information rapidly as soon as it became available and anticipated potential referential interpretations early on. These insights are not new, of course. The present study merely replicates well-known results from studying oculomotor patterns with eye tracking (e.g. Dahan *et al.* 2002, Weber *et al.* 2006, Watson *et al.* 2006, Kurumada *et al.* 2014). Using the mouse tracking method, we showed that when listeners received discourse

relevant intonational information, their hand motions began to curve towards the target response before lexical disambiguation had taken place.

While the literature has mainly looked at intonational processing of rather clear mappings of intonational form and pragmatic interpretation, i.e. the presence vs. absence of a prominent pitch accent indicating contrastiveness, it remains to be seen, how these results generalise to scenarios in which listeners are exposed to more variable intonational information. Intonational categories have been shown to be characterised by a tremendous amount of variability (e.g. Grice *et al.*, in press, *inter alia*), exhibiting no one-to-one mapping of form and function. Future research will need to answer the question as to how listeners accommodate to this degree of uncertainty in intonation-based intention recognition. The present study serves as a proof of concept that such questions can be conveniently studied using the mouse tracking paradigm (see also Tomlinson & Bott 2013, and Warren 2017). We hope that our results spark more interest for this low-cost and pragmatically flexible experimental paradigm for research on speech perception within domains that go beyond phonemic and lexical processing. Mouse tracking proofs to be a fertile method to unravel the real-time dynamics of speech processing such as intonation-based intention recognition.

References

- Bates, D. Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1).
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS ONE*, 5(5), e10773.
- Cangemi, F., Grice, M., & Krüger, M. (2015). Listener-specific perception of speaker-specific production in intonation. In Fuchs, S., Pape, D., Petrone, C., & Perrier, P. (eds.), *Individual Differences in Speech Production and Perception* (pp. 123-145), Frankfurt am Main: Peter Lang.
- CursorSense (2016). (computer software, version 1.3.2). Plentycom Systems. Retrieved from <http://plentycom.jp/en/cursorsense/download.php>
- 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.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, 35(1), 15-28.
- Féry, C., & Kügler, F. (2008) Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 36, 680-703.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226-241
- Grice, M., Niemann, H., Ritter, S., Roettger, T. B. (in press). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*. <https://doi.org/10.1016/j.wocn.2017.03.003>
- Kieslich, P. J., & Henninger, F. (2016). Mousetrap: Mouse-tracking plugins for OpenSesame (Version 1.2.1). doi: 10.5281/zenodo.163404
- Kieslich, P. J., Wulff, D. U., Henninger, F., and Haslbeck, J. M. B. (2017). mousetrap: Process and Analyze Mouse-Tracking Data. R package version 3.0.0. <https://CRAN.R-project.org/package=mousetrap>
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M.K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 133(2), 335-342.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33. <https://CRAN.R-project.org/package=lmerTest>
- Ladd, D.R. (2008). *Intonational phonology*. Cambridge: CUP.
- Magnuson, J.S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, 102, 9995-9996.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314-324.
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). afex: Analysis of Factorial Experiments. R package version 0.16-1. <https://CRAN.R-project.org/package=afex>
- Spivey, M.J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102, 10393-10398.
- Tomlinson, J. & Bott, L. (2013). How intonation constrains pragmatic inference. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 3569-3575.
- van de Vijver, R., & Baer-Henney, D. (2014). Developing biases. *Frontiers in psychology*, 5, 1-8.
- Warren, P. (2017). The interpretation of prosodic variability in the context of accompanying sociophonetic cues. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1), 11.
- Watson, D., Tanenhaus, M.K., & Gunlogson, C. (2008). Interpreting pitch accents in on-line comprehension: H* vs L+H*. *Cognitive Science*, 32, 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, 367-392.