HESITATION PROCESSING ANALYSIS USING CONTINUOUS MOUSE-TRACKING AND GAMIFICATION

Simon Betz[1]*, Éva Székely*[2]*, Sina Zarrieß*[3]*, Marin Schröer*[1]*, Leonie Schade*[1] *and Petra Wagner*[1]

[1]*Universität Bielefeld,* [2]*KTH Stockholm,* [3]*Universität Jena*
simon.betz@uni-bielefeld.de

Abstract: We present a new take on mouse-tracking to analyze online processing of speech. We expand current paradigms by embedding the task into a drag & drop game that awards a score as feedback for performance. The contribution of this paper is twofold: On the one hand, we describe the setup and design of a gamified graphical user interface to elicit data indicating online speech processing in general, on the other hand we provide preliminary results obtained in a small-scale pilot study concerned with the processing of hesitations in speech. Our results serve only to deduce first tendencies which may serve as hypotheses for subsequent analyses, but are sufficient as a proof of concept that our gamified mouse-tracking setup is suitable for studying the online processing of speech.

1 Introduction & background

This paper introduces an experimental setup for analyzing online processing of speech phenomena. The setup is an extension of existing mouse-tracking (MT) and eye-tracking (ET) approaches. Using a graphical user interface (GUI) we engage users in a drag & drop game while listening to audio instructions. We plan to use this setup to analyze online processing of hesitations, such as fillers (*uh, uhm*), lengthening and silences. We report preliminary results from pilot studies, but the main focus of this paper shall be the discussion of the novel methodology.

1.1 Hesitations

Hesitations in speech are phenomena that temporally extend the flow of speech, allowing extra time for speaker and listener. There is a measurable cognitive effect of the use of hesitations. Using ET, it could be shown that listeners are biased towards unfamiliar or new objects when perceiving hesitations [1]. Research into the effect of hesitations predominantly dealt with the two most salient types: silences and fillers. The third major type of hesitation, lengthening, is under-researched in this respect. In our own previous studies, we found that lengthening is a very elusive phenomenon in the speech signal, which can be reliably spotted with semi-automatic classification, but which often passes unnoticed by human listeners [2]. With the experimental paradigm presented here, we thus hope to show that there are differences in processing between lengthening and fillers, which would demand a more differentiated view on the communicative effect of hesitations.

1.2 Mouse tracking

In recent years MT has been used in a variety of studies across different fields to gain insight into the underlying cognitive processes in forced choice tasks and to answer the question if

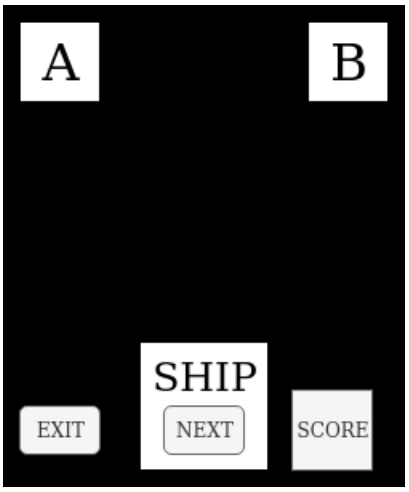


Figure 1 – Layout of graphical user interface. A and B are the target zones, SHIP is the draggable object, SCORE displays the game score, NEXT button starts a new trial, EXIT closes the application.

these processes are dynamical or discrete (e.g. [3], [4]). Usually the participants are given a choice between two competing objects that are located in the upper corners of the screen and are instructed to move the cursor as quickly and accurately towards the target object while a stimulus is being presented. From the resulting cursor path one can infer the initially preferred competitor as well as if there has been a reanalysis. Mouse paths can be analyzed by measuring the resulting `AREA UNDER THE CURVE (AUC)` which will increase when there is a bias from a competing object. In general, MT is more user-friendly to handle compared to ET. Participants can be seated at a PC and start the experiment without any invasive technology attached and without individual calibration necessary. However, MT also poses some difficulties (see also [5] for an overview).

1.3 Aim

The first goal of this paper is the exploration of the potential of MT enriched by aspects of gamification. In order to obtain insights about the online processing of hesitations, we engage participants in a drag & drop task while audio stimuli are presented. This way we expand the traditional MT paradigm by the explicit inclusion of the movement along the path into the task. We refer to this concept as `continuous MT` to differentiate from the typical `click A or B` setup. The second goal of this paper is the preliminary exploration of a research question, based on the observation by Arnold et al. [1] that fillers bias listeners towards unfamiliar objects. From this, as well as our previous findings on the perceptual subtlety of hesitations expressed by lengthening, we derive two exploratory hypotheses: (H1) AUC increases if participants are instructed to move towards a concrete, familiar, target object while listening to a stimulus containing a hesitation with a filler. Furthermore (H2), we expect that due to their perceptual subtlety, stimuli containing hesitations marked by lengthening rather than fillers result in a different behavior.

2 Method

2.1 The GUI

The basic layout of the GUI is a traditional MT setup, with target zones in the top left and top right corners of a portrait-format screen, cf. Fig.1. In addition, there is a draggable object (SHIP) in the bottom middle of the screen. As the experiment is devised as a game, the entire design has a space theme to give a background. In the bottom right corner, a game score can

be displayed to live comment on the performance. On the ship there is also the NEXT button, which initiates a new trial, playing the next audio stimulus, creates new objects in A and B and relocates the ship to the starting position.

2.2 Technical specifications

The experiment was coded in python using the tkinter package and run on a laptop with Ubuntu 19.4 as operating system. Using gnome tweaks, the mouse speed was adjusted so that it was not straightforwardly possible to reach the opposite end of the screen before the audio finished. The mouse used was a Pulsefire FPS Pro gaming mouse with 16.000+ dpi resolution, used on a steelseries QCK+ gaming mouse pad sized 40x45cm.

2.3 Stimuli

Following the idea of Arnold et al. [1], there are `CONCRETE` and `ABSTRACT`[1] target objects in order to examine whether the hesitation bias towards unfamiliar objects can be replicated. The concrete objects are images of planets, satellites and space monsters. The abstract objects are three different random shapes combined with nonsense names (*Bogron, Sombrum, Diliar*). All objects can appear in any of three colors: red, blue and green. The corresponding audio stimuli follow a regular structure of "`FLY TO THE <COLOR> <OPTIONAL HESITATION> <SHAPE>`". The audio comes in three conditions, `NO` (baseline with no hesitation), `LEN` (hesitation lengthening) and `FULL` (lengthening and filler). This results in 54 combinations of colors, shapes and hesitation conditions. All stimuli were presented twice, yielding 108 stimuli in total. The stimuli were created by recording a speaker who was instructed to produce each combination once and utter a filler (*uhm*) after the color word. Based on this, condition `FULL` was created by lengthening the final nasal in the color word by about 500ms using phonetics software (Praat), resulting in lengthening + filler. Condition `LEN` is the `FULL` condition with fillers excised. Condition `NO` is the initial recording with fillers excised. Due to the grammatical structure of German, each color word gets the ending *-en* in the Dative case, providing a nasal sound, which is the optimal target for lengthening [6]. Lengthening of the final sounds in a property word followed by an object word is associated with uncertainty about the following object [7], which should be reflected in a bias towards `ABSTRACT` objects. The GUI creates a shuffled list of audio stimuli when invoked. In each trial, the visual target object is read from the audio file name and is assigned randomly to position A or B. Correspondingly, a competitor object is created in the other position. The competitor is chosen randomly among candidates with same color and opposing abstractness.

2.4 Game score

As is the case in other MT experiments, participants are instructed to move as precisely and quickly as possible. In this setup, this poses challenges, as it is crucial that the movement is executed while the audio stimulus is playing. In a first testing phase it became apparent, that participants would frequently try to break the system by choosing paths that are detrimental for analysis. For that reason, we included a game score that would reward movements in a given time frame and a given precision, while punishing deviations. We set a base score of 500, which would be increased by 250 if the participants finished the movement within a 400ms window around the audio stimulus end. Slower movements would be punished by subtracting the elapsed time in ms from the score, in order to prevent participants from waiting for the end

[1] "abstract" in the sense of "hard to describe, lacking a name in the real world, resembling abstract art".

of the instruction before initiating the movement. Faster movements would be punished by a fixed score reduction in order to prevent 50% chance guessing tactics.

2.5 Participants & training

For this pilot study, $n = 8$ participants took part in the experiment. They received no monetary compensation but could leave their email addresses to have a chance of winning a 15 Euro voucher in case they reached the highest score. We collected information about gender, age, handedness, mother tongue, eyesight and hearing, as well as asking for experiences with several mouse-based video games. Participants were briefed about the functionality of the GUI and how the game score is computed. They were given the hint that a more or less continuous path towards the target is a good way to reach a high score. Prior to the experiment there was an open training phase under supervision of the experimenter, in which participants were allowed to do as many practice trials as they wanted to familiarize with the game objects and the unusually slow cursor speed.

2.6 Descriptive statistics

We hypothesized to see an impact of a filler + lengthening (`hesitation = FULL`) hesitation present in the stimulus on the participants' movement behaviors, leading to a bias towards the competing target object and hence to higher values of `AUC` in cases where the target concept is concrete (`ABCO = concrete`). However, for the more subtle sounding hesitation expressed purely by lengthening (`hesitation = LEN`), we did not expect this effect. Given the likelihood of subtle effects and data sparsity in this pilot study, we restrict ourselves to reporting descriptive statistics and trends, focusing on potential effects that could serve as hypotheses in subsequent experiments. More specifically, we calculated means and standard deviations of the absolute measure of `AUC` across different participants and under the factor `hesitation`, comprising three factor levels (`NO, LEN, FULL`), and the factor `ABCO` comprising two factor levels, denoting either `concrete` or `abstract` target concepts. For this pilot study, we obtain only raw `AUC` measures. For a discussion of finer-grained analysis methods, see [8].

3 Results & discussion

All apparent trends are subtle and need to be interpreted with caution. Both types of lengthening (`LEN, FULL`) led to a slight increase in `AUC` compared to the baseline condition not containing any lengthening (cf. Table 1). It is likely that this is an artifact of the study design, as the lengthened stimuli gave participants more time to move, and hence, they had more opportunity to move away from the target. There is an overall slightly larger `AUC` for abstract target objects (cf. Table 2). This points towards a slight general bias for preferring the more concrete target objects, and a similar trend can be found in 6 out of 8 participants (cf. Figure 3). In general, it has to be considered that participants may pursue individual strategies during a game. Fig. 4 illustrates the mouse movements of the first three participants, showing different degrees of space coverage and directness of motion.

As we expected an interaction effect, we looked at the impact of `hesitation` for the two different target types, `abstract` vs. `concrete` (cf. Figure 2). The mean values indeed indicate a potential interaction effect which at least partly supports our expectations: If the target objects are of type `abstract`, the hesitation condition `FULL` indeed leads to a slight increase in the `AUC`. In the baseline hesitation condition `NO` we find the slightly lower value for the concrete target objects reported above. The hesitation condition `LEN` leads to a smaller `AUC` for concrete target objects as well, but also shows a considerably larger `AUC` for abstract targets. Taken

Hesitation type	AUC, mean (sd)	n
NO	15.99 (12.1)	287
LEN	17.1 (12.53)	286
FULL	16.7 (11.65)	287

Table 1 – `AUC (abs)` for different hesitation conditions `NO`= 'no hesitation', `LEN`='hesitation lengthening' and `FULL`='lengthening + filler'.

ABCO	AUC, mean (sd)	n
abstract	16.85 (12)	432
concrete	16.29 (12.2)	429

Table 2 – `AUC (abs)` for abstract vs. concrete target objects.

together, our data indicate a preference bias for the concrete stimuli across most participants, which is strengthened in the presence of hesitations consisting of only lengthening (`LEN`). In the presence of hesitations consisting of lengthening + fillers (`FULL`), this bias appears to be overridden, leading to a preference for the abstract object, which would confirm the hesitation bias for unfamiliar objects caused by fillers. It needs to be determined in follow-up studies whether the tendency of lengthening behaving differently than fillers can be confirmed with a larger dataset.

4 General discussion & outlook

While the results regarding the research question can only be interpreted very carefully given the exploratory character of this study, there are several insights regarding the methodology gained in this study, and several open questions are up for discussion. Generally, it can be concluded that the MT and gamification approach works satisfactorily in terms of producing meaningful and analyzable results. The inclusion of a game score seems to motivate participants to maximally adhere to the task given, and the playful environment prevents fatigue which is often troubling in monotonous linguistics experiments. However, there are areas that need improvement or that need to be added for the continuation of this series of studies:

- **Score logging**. At the moment we use the game score only as inherent motivation. Logging the score per run would enable to obtain more precise insights into participant behavior and account for inherent difficulty of runs.

- **Evaluate precision**. Currently, precision only feeds the game score and is not evaluated further. It would be desirable to have precision metrics on top of score logging for finer grained assessment of the results.

- **Advanced measurements**. In addition to raw AUC measures, the cursor path around the moment of hesitation would be of interest in order to monitor cognitive processes. Also, further analysis methods, as described in [8] could be considered for future work.

- **Fuel efficiency**. Participants with gaming background figured out that it is still possible, despite the game score design, to rush to the top middle of the screen, wait for the audio to finish, and rush over to the target. This behavior produces paths which are not telling for the research question at hand. It could be an option to include fuel efficiency as a factor into the game score and display it online as a function of distance covered.

- **Stimulus length issue**. Including different degrees of hesitation always poses the problem of comparability. For this game setting, it would be desirable to have stimuli of equal

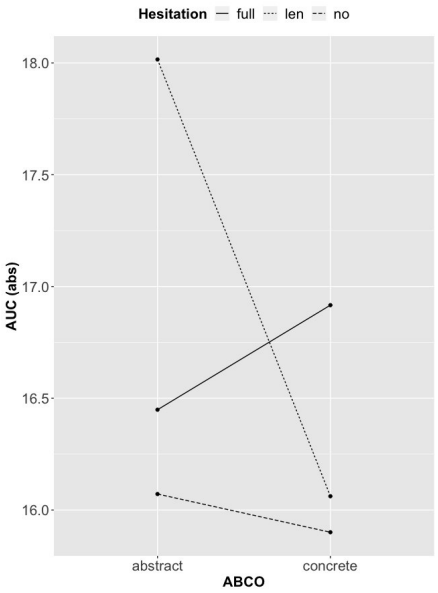


Figure 2 – Potential interactions between `Hesitation Condition` and `ABCO` on participants' `AUC`.

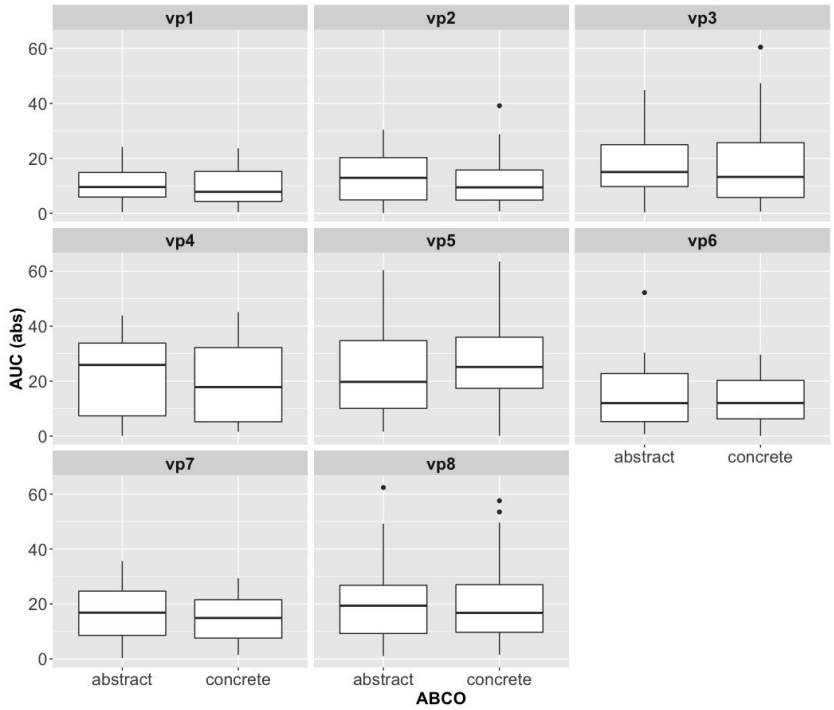
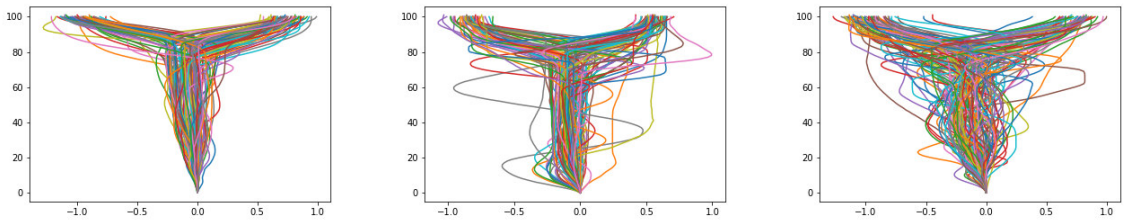


Figure 3 – Individual participants' (vp1—vp8) `AUC` as a function of concrete and abstract target objects.



(vp1) (vp2) (vp3)

Figure 4 – Participants vp1-vp3's mouse paths. Y-axis shows progress normalized to 100 steps, X-axis is the X-position of the ship, normalized to 0 being the starting point.

duration. This, however, could only be achieved by manipulating speech rate or speech material in the NO and LEN conditions, which have inherently shorter duration. A possible solution would be to add noise, e.g. rocket launch sounds to the beginning of the stimuli to disguise slightly differing starting points. The beginnings of the stimuli could overlap with noise, as they are similar across all runs. This would enable to have the critical stimulus ends aligned, which contain the relevant information.[2]

- **The voice**. Originally, this study was devised for speech synthesis. As there were notorious problems synthesizing fillers, we replaced it by a recorded voice. The quality gain comes with the problem that the fillers must be "acted" in order to have control over them. It would be possible to elicit hesitations with a naming task (cf. [9]), which, alas, would detriment controlability, given that every speaker has a different way of producing hesitations [10]. For future work it is thus necessary to motivate the choice of voice thoroughly.

References

[1] ARNOLD, J. E., C. L. H. KAM, and M. K. TANENHAUS: *If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), p. 914, 2007.

[2] BETZ, S., J. VOSSE, S. ZARRIESS, and P. WAGNER: *Increasing recall of lengthening detection via semi-automatic classification*. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (Interspeech 2017, Stockholm)*, pp. 1084–1088. 2017.

[3] FREEMAN, J. B. and N. AMBADY: *Motions of the hand expose the partial and parallel activation of stereotypes*. *Psychological Science*, 20(10), pp. 1183–1188, 2009.

[4] ROETTGER, T. B. and M. FRANKE: *Evidential strength of intonational cues and rational adaptation to (un-) reliable intonation*. *Cognitive science*, 43(7), p. e12745, 2019.

[5] GRAGE, T., M. SCHOEMANN, P. KIESLICH, and S. SCHERBAUM: *Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition*. *Attention, Perception, Psychophysics*, 81, 2019. doi:10.3758/s13414-019-01889-z.

[6] BETZ, S., P. WAGNER, and J. VOSSE: *Deriving a strategy for synthesizing lengthening disfluencies based on spontaneous conversational speech data*. In *Phonetik und Phonologie 12*. 2016.

[7] BETZ, S., S. ZARRIESS, E. SZÉKELY, and P. WAGNER: *The greennn tree-lengthening position influences uncertainty perception*. In *Proceedings of Interspeech*, pp. 3990–3994. 2019.

[8] HEHMAN, E., R. M. STOLIER, and J. B. FREEMAN: *Advanced mouse-tracking analytic techniques for enhancing psychological science*. *Group Processes & Intergroup Relations*, 18(3), pp. 384–401, 2015.

[2]Thanks to Dennis Hoffmann for the inspirations regarding fuel efficiency and rocket noise.

[9] BRENNAN, S. E. and M. F. SCHOBER: *How listeners compensate for disfluencies in spontaneous speech*. *Journal of Memory and Language*, 44(2), pp. 274–296, 2001.

[10] BETZ, S. and S. LÓPEZ-GAMBINO: *Are we all disfluent in our own special way...and should dialogue systems also be?* In O. JOKISCH (ed.), *Elektronische Sprachsignalverarbeitung (ESSV) 2016*. TUD Press, 2016.