

University of Groningen

Salience is in the eye of the beholder

Boswijk, Vincent H.; Loerts, Hanneke; Hilton, Nanna Haug

Published in:
 Ampersand

DOI:
[10.1016/j.amper.2020.100061](https://doi.org/10.1016/j.amper.2020.100061)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Boswijk, V. H., Loerts, H., & Hilton, N. H. (2020). Salience is in the eye of the beholder: Increased pupil size reflects acoustically salient variables. *Ampersand*, 7, [100061].
<https://doi.org/10.1016/j.amper.2020.100061>

Copyright

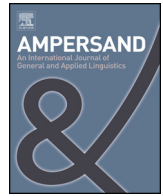
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Saliency is in the eye of the beholder: Increased pupil size reflects acoustically salient variables



Vincent Boswijk*, Hanneke Loerts, Nanna Haug Hilton

University of Groningen, the Netherlands

ARTICLE INFO

Keywords:

Saliency
Speech perception
Pupil dilation
Processing load
Generalized additive mixed modelling (GAMM)

ABSTRACT

'Saliency' is a term frequently used in linguistics but an exact definition for the concept is lacking. Recent technological advances which allow us to explore the cognitive processing of so-called salient linguistic features could provide us with quantifiable measures of 'saliency', and lead to a further understanding of the concept and its relationship to language acquisition and change. In this paper we measure pupil dilation with the assumption that auditory saliency results in a change in pupil size, as an effect of cognitive load. We report an experimental study observing Dutch participants' pupil sizes when listening to stimuli containing salient and non-salient variants of linguistic variables (e.g. Dutch coda/r/; speech intensity, word frequency). Using Generalized Additive Mixed Modelling (GAMM), we find pupil size increases for three of six stimuli categories. We consider our findings in light of the speech processing literature, address the (dis)advantages of the technique, and formulate some recommendations for future advances in neurophysiological measures in (socio)linguistics.

1. Introduction

'A maddeningly underdefined term' is Meyerhoff's [1] definition of the term 'saliency'. While there is common ground between different notions of what linguistic saliency entails, the exact meaning of the term is seen as difficult to define [2]. The concept of saliency assumes some form of psychological prominence [3], and attempts to combine both structural (language-internal) factors with sociolinguistic and psychological (extra-linguistic) factors in a single explanatory concept [3]. However, none of the existing definitions cover all of the aspects of saliency that are currently in use. Furthermore, many examples where saliency is operationalized involve circular logic [3,4], in that the properties of a salient variable are those that follow from it being salient in the first place [4] (e.g. the operationalization of Trudgill [5]). Thus, the use of the concept oftentimes raises more questions than it answers and one could question whether 'saliency' has explanatory value at all [4].

This paper reports on a study of whether pupil dilation can be used as a quantifiable measure of saliency by testing a neurophysiological response, pupil-dilation, to six different operationalizations of the term 'saliency'. The aim of our paper is to evaluate this particular methodology of quantification. The theoretical importance of a tool that can quantify saliency could be substantial for fields such as sociolinguistics and language acquisition. In sociolinguistic studies of language change,

for instance, salient features are believed to receive more attention, and are thus generally (albeit not always) accommodated towards more easily than other features [3]. Deconstructing the notion of saliency, then, seems crucial to determining whether the concept is a fruitful predictor for language change.

As such, the aim of this paper is not to establish an overarching working definition of saliency, but rather to experimentally investigate and compare existing definitions. By investigating how different 'salient' traits are perceived, we aim to find out whether there is indeed reason to call them thus. Although the topic of interest is linguistics, we will see that some operationalizations are not limited to linguistics, but are also potentially of interest in other domains.

2. Saliency in linguistics

The origin of the concept of 'saliency' in linguistics lies in the work on dialect contact in German speaking enclaves in Russia by Schirmunski [6]. Introducing '*auffälligkeit*', Schirmunski distinguishes between primary (cf. salient) variables, which are susceptible to change and loss, and secondary (cf. non-salient) variables, which are stable. Since then, many have tried to define the concept of saliency through similar sets of linguistic criteria, or through experimental studies. Nowadays, 'saliency' is found throughout linguistic subdomains, and although definitions overlap, there is no all-encompassing definition of

* Corresponding author. Wirdumerdijk 34, 8911, CE, Leeuwarden, the Netherlands.

E-mail address: v.h.boswijk@rug.nl (V. Boswijk).

the concept that subsumes usage across linguistic subfields.

Linguistic traits such as loudness, high word-frequency, or a greater articulatory effort have been put forward as ‘salient’, whereas it has also been argued that salience is a result of associations with social factors [3]. In semantics and morphology, for example, salient features are the more regular or higher-frequency constructions. According to the operationalization of Giora [7] the salient meaning, which can be said to be ‘foremost on one’s mind’, is the more accessible one, and it is thus easier to process. In phonology, on the other hand, salience is a feature of the more irregular or lower-frequency constructions. It is said that these salient features are more prone to attract attention, and standing out more than others.

It has to be noted that salience is a relative concept [4], and it should as such always be seen in the appropriate context. This holds true for the linguistic context, i.e. how probable a variant is given its phonological and lexical surroundings, and the social context, i.e. how probable it is based on socio-indexical information about the person speaking [8]. Furthermore, the degree of salience might be influenced by the current tasks and goals of the perceiver [9].

A recurring theme within the literature on ‘salience’ is that it is often linked to cognitive processes dealing with attention or processing difficulty [10,11]. Correlates of cognitive processing can be measured using neurophysiological measures, such as functional Magnetic Resonance Imaging (fMRI),¹ or Event Related Potentials (ERPs),² but also with less costly and more mobile measures such as pupillometry, which measures pupil reactivity based on pupil size. Liao et al. [12] suggest, for example, that pupil dilation reflects cognitive functions related to attention and salient stimulus detection, and numerous other examples of experimental findings point towards a relationship between cognitive processes and pupil dilation [13–16] (see section 4 for a more detailed discussion of pupillometry). This means that we can use neurophysiological measures to quantify the cognitive processes that take place when we are exposed to linguistic features that stand out. One can thus argue that pupillometry could be a relatively cheap and mobile measure of the degree of salience of a linguistic feature.

Ultimately, the objective measure of responses to salient features can lead to new possibilities in monitoring the perception of language change. Understanding how the perception of specific linguistic features relates to changes in language production can lead us closer to addressing theoretical concerns in linguistics, such as the ‘actuation problem’ [17] in sociolinguistics (i.e. how and why language change starts out).

3. Operationalizations of salience

One of the problems that surrounds the concept of salience, as illustrated by Llamas, Watt and MacFarlane [18], is the inconsistency and the apparent arbitrariness of the properties that make a feature salient. The stimuli we test in our experiment are based on examples from the literature. These various properties of language have been considered salient in earlier work. Some of these properties are more likely to be salient cross-linguistically, while others are only salient in specific contexts or in a specific speech community at a particular point in time [19]. These properties will be discussed below.

Liao et al. [12], for example, discuss the relationship between the salience of sounds and their loudness. They find that auditory salience can be characterized by contrasts in stimulus characteristics, such as

¹ In fMRI, brain activity is mapped by measuring the blood flow to different parts of the brain. The data collected through fMRI is particularly suitable for providing information about the location of brain activation.

² Brain activity can be measured by recording an Electroencephalogram (EEG). Different tasks and/or stimuli elicit specific time-locked responses in the EEG signal known as ERPs. These ERPs are especially informative when investigating the time-course of processing.

intensity. Although, so they state, it is known that deviant or contrasting auditory stimuli evoke an increase in pupil size, a link with pupil dilation and the subjective salience of sounds had yet to be made. In their eye-tracking study, Liao et al. [12] seek to do just this by linking the results of a subjective judgment task concerning characteristics such as salience, preference, loudness and beauty, to pupillary responses. Their participants listened to random sound pairs and were asked to determine which stimulus was more beautiful, noticeable, remarkable, etc. Not only did they find that pupil size significantly increased for stimuli that had greater intensity, they also found that these stimuli were deemed more salient in a subjective judgment task. These results suggest a close link between salience and loudness, and additionally show that pupil dilation is a likely predictor of a feature’s perceived salience (in this particular operationalization of the term).

Furthermore, contrasts in phonology have been hypothesized to be salient (e.g. [5]). According to Hickey [2] the realization of a phoneme with a previously unheard variant makes a feature salient due to an unexpected phonological distinction. He calls this distinction acoustic prominence. This criterion follows from Schirmunski [6] who states that features that are phonetically different tend to be salient in their surroundings.

A third feature that has been considered as salient is the use of non-standard grammatical features. Many such examples exist, such as was/were regularization or negative concord in British English (cf [3]. for more examples). For Dutch, the realization of grammatical gender is a linguistic variable with a clear standard and non-standard (often perceived as entirely ungrammatical by a part of the population) variant [20]. The usage of the non-standard variant of grammatical gender in Dutch has been reported to be salient by, amongst others, Hanulíková et al. [19]. Grammatical gender is thus a suitable feature for testing the processing of salient variants in Dutch.

Moreover, several studies discuss the relationship between the relative frequency of a linguistic feature and its salience. This relationship between frequency and salience is not without discussion. Some [7,21,22] state that a feature with a higher linguistic frequency is more salient because it is ‘more forward on one’s mind’ [7] and hence easier to process. Such a notion is in line with experimental studies finding a relationship between an increase in frequency and ease of processing (e.g. Ref. [23]). On the other hand, scholars like Zarcone et al. [9], argue that linguistic features that have a low frequency are more salient, in that these stand out because they are harder to predict in regard to the probabilistic expectations of the upcoming signal. This is often linked to the possible relationship between salience and surprisal. Interestingly, like salience, surprisal can be linked to (implicit) learning and processing difficulty [8]. Furthermore, surprisal plays an important role in the adaptation of expectations, which in turn is important for speech perception [8]. Importantly, “surprisal and salience both affect language processing at different levels, but the relationship between the two has not been adequately elucidated, and the question whether salience can be reduced to surprisal/predictability is still open” [9].

Paradoxically, then, both high-frequent and low-frequent linguistic features may be labelled as particularly ‘salient’. In our experiment we compare responses to high-frequency features with responses to features with low-frequency rates.

Sociolinguistic studies often include examples of features that are salient due to non-linguistic, social, properties (e.g. markers of ingroup and outgroup status, see Ref. [18]). Thus, features may acquire social meaning based on certain social characteristics attached to those speakers that use them. Speakers in turn may then adjust their use in line with these characteristics in order to signal social information to other speakers [4,24]. These features are called ‘markers’, or even ‘stereotypes’ when they become overtly stigmatized, and are likely to evoke value judgements: speakers who use certain linguistic features, e.g. h-dropping in British English, can be associated with particular social stigmas. Such features, then, have ‘salience’ purely due to awareness of their existence on the part of members of the speech

community of their existence. An example from Dutch of a feature with a high level of awareness is the retroflex bunched approximant pronunciation of /r/: the variant more commonly known as ‘Gooise r’ [25], which is reported to have ‘relatively high sociolinguistic salience’ [26]. The variant strongly resembles the bunched approximant pronunciation of /r/ in varieties of American English, and can only occur post-vocally, such as vowel + r + # in words like *hoor* (hear) and *weer* (weather) and vowel + r + consonant(s) + # in words like *bord* (plate) and *dorst* (thirst) [25]. Van Bezooijen and van den Berg [25] illustrate the level of awareness towards the feature with a number of examples from the Dutch media, which show that the variant regularly comes with negative associations and is often met with irritation. One such example discusses how the variant featured in the famous Dutch television program *Man bijt Hond*, in the category ‘what is getting on your nerves?’ as the ‘pompous r’. This feature is thus highly suited to testing the processing of variables in Dutch of which speakers are aware. To consider whether this operationalization of salience can be measured quantitatively with neurophysiological measures, a linguistic feature of which people are unaware ought to also be used in the experimental design.

There exist features that are part of language variation and change, and that may even be reflective of a person's social or regional background, that we are not aware of. These are generally referred to as ‘indicators’ [24], and changes that take place in such features are generally below the level of consciousness. As an example of such an indicator, Rácz [27] mentions the distinct pronunciations of words like *what* and *where* in Northern and Southern American English. Although Southern dialects may use the velar approximant, while Northern dialects do not this characteristic does not seem to be identified as being typically Southern [27]. According to the sociolinguistic definition of salience, indicators should never be salient. An example of a recent sound change in Dutch below the level of consciousness, is the pronunciation of word-initial fricatives as voiceless [28]. The pronunciation of /v/ as either voiced [v] or voiceless [f] is a variable realization speakers are mostly unaware of [28]. This particular variable, then, lends itself to testing the processing cost of indicators in Dutch.

4. Pupillometry and cognition

Generally, pupil size varies between 2 and 8 mm, and under normal circumstances the pupil is about 3 mm in diameter [29]. Van Rijn, Dalenberg, Borst and Sprenger [30] report that pupil dilation is a relatively slow measure, and indeed, the light reflex has for example been reported to occur with a latency of around 220 ms [31]. Variations in pupil size based on cognition are small fluctuations in size which are usually smaller than 0.5 mm [32].

Interest in pupil size as a measure for cognitive processes originated, according to Hyönä, Tommola and Alaja [14], in the 1970s. Interest was sparked by the works of Hess and Polt [33], who have shown that pupil size was related to cognitive processing demands, and Kahneman and Beatty [34], who were able to show that a larger amount of material in a short-term memory task was associated with a larger pupil size.

Modern examples of studies using eye-tracking to examine pupil dilation and cognition include Liao et al. [12], who found a correlation between loudness and pupil dilation, Koelewijn, de Kluiver, Shinn-Cunningham, Zekveld and Kramer [15], who pointed out that a higher level of listening effort can be identified through an increase in pupil size, Vogelzang, Hendriks and van Rijn [35], who found that in pronoun processing, increased pupil size reflects greater difficulty in ambiguity resolution, or Mathôt, Grainger and Strijkers [36] who found that pupil size is altered by words that convey a sense of darkness or brightness. Other studies using pupillometry have successfully shown that pupil dilation is associated with, amongst other things, emotional arousal [37], memory load in word retrieval [13], decision making [38] and word frequency [39].

Beatty and Lucero-Wagoner [32] add that the processes associated with pupil dilation ‘reflect variations in central processing load with extraordinary precision’. Hence, the use of pupillometry in the study of language processing cost has been called ‘remarkably consistent and without significant contradictions’ [32].

Furthermore, the aforementioned examples tell us that there is a relationship between pupil dilation and cognitive load. In the words of Paas, Tuovinen, Tabbers and van Gerven [16], cognitive load represents ‘the load that performing a particular task imposes on the [...] cognitive system’. Hyönä et al. [14] conclude that pupillary responses can be used as a reliable moment-to-moment measure of processing load. One key challenge of the use of pupillometry, as with most physiological measures, is multi-causality. As demonstrated above, there are a number of possible reasons for pupils to dilate, and it is hard to exclude potential confounding causes from the experimental manipulations [40].

5. The Present Study

The primary research question for our experiment is: *does the processing of salient linguistic features result in an increase in pupil diameter?* In other words, can salience be measured in pupil size? For this study, we measure pupil size while presenting participants with auditory salient stimuli. Based on the review above, we hypothesize that if salience is related to cognitive processes such as attention or cognitive load this has to be visible as an increase in pupil size during the processing of spoken language. To find out whether this was indeed the case, we asked participants to listen to speech samples with manipulated loudness, articulatory prominence, and the use of non-standard grammatical features, as well as the use of the innovative/r/-variant. We predict that Dutch listeners will experience more difficulties in speech processing when presented with the features that are hypothesized to be salient, than when listening to the regular stimuli. We expect that this greater difficulty will result in an increase in pupil size.

As discussed above, a unified definition of salience within linguistics is lacking. Accordingly, this experiment, bases itself upon several operationalizations from the literature. By testing pupillary responses to these different operationalizations, we aim to find out whether there is overlap between the cognitive responses to these different operationalizations. By comparing the responses we might then further our understanding of the concept of salience within linguistics. This could potentially help us to unify different definitions of ‘salience’.

Based on the literature on salience we predict that Dutch listeners hearing speech samples with manipulated loudness, articulatory prominence, and the use of non-standard grammatical features, as well as the use of the innovative/r/-variant will experience more difficulties in speech processing than when listening to samples with a regular speech volume, standard articulatory patterns, standard grammatical features, and the use of the traditional r-variant. We further expect that frequency will affect processing load, but we make no hypotheses about the direction of the effect. Finally, the use of a variant in a sound change below the level of consciousness is included to test the hypothesis that this will not affect processing difficulty.

6. Methods

To test our hypothesis, we conducted a twofold experiment. In the first, quantitative, part of the experiment, participants listened to variables from six different categories of salience (see Table 1) while their pupil size was monitored. In a second, qualitative part, we asked the participants a number of questions to find out whether they had indeed perceived our hypothesized salient stimuli as such. Because of the possible interference of pre-exposure to the stimuli, it was important to follow this specific order.

Table 1
The categories and their conditions.

Category	More likely to be salient	Less likely to be salient
Loudness	Word intensity of 80 dB	Word intensity of 60 dB
Acoustic Prominence	/t/pronounced as [p]	/t/pronounced as [t]
Grammatical Gender	Violation of grammatical gender	Correct grammatical gender
Frequency	High frequency word	Low frequency word
Conscious Sound Change	/r/pronounced as [ɹ]	/r/pronounced as [r]
Subconscious Sound Change	/v/pronounced as [v] (=voiced)	/v/pronounced as [f] (=voiceless)

6.1. Stimuli

Stimuli consisted of spoken samples created on the basis of the notions presented in section 2 and divided over six categories that we called *Loudness*, *Acoustic Prominence*, *Grammatical Gender*, *Frequency*, *Conscious Sound Change*, and *Subconscious Sound Change* (Table 1). Each category consisted of 8 salient – non-salient sample pairs. Each sample pair shared the same carrier sentence in which the variable was embedded.

The stimuli chosen for the category Loudness were based on a study by Liao et al. [12], in which loudness was shown to influence subjective salience, which in turn was related to increased pupil size. This category thus served as a control in our experiment. The salient stimuli had an altered intensity of 80 dB, as opposed to an intensity of 60 dB for the carrier sentences and non-salient stimuli.

For Acoustic Prominence, participants were presented with a previously unheard variant as described by Hickey [2]. Thus, we used stimuli in which /t/ was either realized as [t], or as [p], the latter of which was hypothesized to be salient. This was done in such a way that the stimuli formed non-existing words. As such, it was impossible that the stimuli with [p] could be seen as an acceptable realization of a different word.

The Grammatical Gender category presented the participants with both correct and incorrect usage of grammatical gender in the definite article in Dutch. The incorrect forms were hypothesized to be salient, whereas the correct forms were expected to be non-salient, which is in line with, among others, Hanulíková et al. [19]. An interesting finding is that this is not the case if the speaker is expected to make gender violations, showing the important role of context [19]. The Dutch language has two possible genders: around 75% of nouns have common gender, receiving the definite article ‘de’, and around 25% have neuter gender, receiving the definite article ‘het’ [41]. In the past, incorrect use of grammatical gender by a native speaker has been shown to invoke online repair processes [42].

For Frequency, we did not make explicit hypotheses about the direction of the effect (as discussed above). However, for the sake of the experiment, the variables with high frequency were called salient, and the low frequency variables were termed non-salient. The words in the frequency category were taken from a list of fairly similar high-frequency – low-frequency word pairs presented by Rommers [43]. We controlled for word frequency in the other categories, so that the possible effect of frequency would not interfere. We established frequency by using three corpora: *A Frequency Dictionary of Dutch* [44], *Corpus Gesproken Nederlands* [45] and *Twitter Ngrams* [46]. Based on these three sources, mean frequency was calculated and words were selected in such a way that the mean frequency did not differ significantly between the salient and non-salient conditions throughout the categories other than the Frequency category.

In the category Conscious Sound Change, the hypothesized salient variables presented the listeners with the use of the variant ‘Gooise r’, an approximant realization of postvocalic /r/, which has previously been pointed out as being salient by Sebregts [26]. The non-salient variant was the alveolar tap, which is the variant most frequently used

in the north of the Netherlands, where this experiment was conducted [47].

Finally, the stimuli for Subconscious Sound Change consisted of voiced and voiceless realizations of fricative /v/ in Dutch. As described by Pinget [28] the voiceless realization is gaining ground, but is below the level of consciousness. If salience is a feature of conscious awareness, we expect there to be no pupil dilation for the voiceless realization of /v/. However, if lack of a phonetic contrast between voiced and voiceless features is salient, there will be a pupil dilation effect between the two variants [f] and [v].

For all categories we considered possible memory-effects on pupil dilation [c.f. 29]. Because of this, all words and carrier sentences occurred only once for each participant. This way, recognition of the words and sentences could not interfere with the results.

The samples were recorded in a sound studio using Adobe Audition CS6, with the help of a 26-year old female mother tongue speaker of Dutch. The samples were analyzed and edited using Praat [48] and spliced in such a way that each sample pair differed only at the variable level. An overview of the stimuli per category and their transcriptions can be found in Appendix A.

6.2. Design and procedure

The spoken samples were presented using E-prime 2.0 [49] and the E-prime extensions for the Tobii eye-tracker, TET [50]. During the experiment, participants listened to the samples via headphones while pupil data were collected using a Tobii T120 eye tracker with a sampling rate of 60 Hz [51]. The qualitative part of the experiment was recorded using an Olympus Digital Voice recorder WS-200S.

Typically, pupillary responses are measured against a baseline which is recorded before the presentation of the task in order to find the relative increase in pupil size [15,32,35]. We also use such a baseline in our experiment. Participants were seated behind the eye-tracker and asked to move and blink as little as possible during the experiment. After a short explanation of the procedure, a fixation cross would appear in the center of the screen that participants were asked to focus on. The fixation was followed by the auditory stimulus, which was then followed by a screen with asterisks (****). When this screen was shown, the participants were allowed to blink freely. The auditory stimuli were presented in a randomized order, and responses were measured during a 5000 ms period starting from the onset of each sound file. For each stimulus, we made sure that half of the participants listened to the salient variant, while the other half listened to the non-salient variant. In order to do so, we created two versions of the experiment (see appendix A). Participants listened to either version 1 or 2 of the experiment. We then compared results between conditions.

In order to keep participants focused during the experiment, they were asked whether or not they had just heard a certain word, at random intervals. They could answer by pressing a button. Participants were instructed beforehand that questions would be asked throughout the experiment.

The qualitative part of the experiment presented the participants with a stimulus pair from each category, in which one variant was more likely to be salient than the other. Participants could control the audio themselves and, as such, were able to listen to both variants (i.e. the salient and non-salient variant of each stimulus) multiple times. For each category, the participants were asked to elaborate on what they heard. To this end, they were asked if there was something that they specifically noticed, and if so, what it was that was noticeable and why. They were then asked what they thought of this feature. Presentation of the different categories occurred in five different orders, resulting in 5 people listening to each order of presentation (with the exception of one order, which was presented to 6 participants).

6.3. Participants

In order to evaluate our hypothesis, we tested a total of 41 participants (25 females) with a mean age of 23.05, ranging from 18 to 29 years old. All participants were currently enrolled in, or had completed a form of higher education. Participants were randomly divided over the two versions of the experiment. Eight males and thirteen females listened to version 1. Their mean age was 23 years old. Eight males and twelve females listened to version 2. Their mean age was 23.1.

All participants were mother tongue speakers of Dutch. Further language background was collected via a questionnaire. Before testing, participants were asked to fill out this questionnaire, which consisted of questions about their gender, age and educational background, as well as information about their places of residence in order to control for dialectal background.

6.4. Data processing

The data files from E-prime were merged for all participants and uploaded to the statistical environment R [52]. The data set was then cleaned and pre-processed in R. This included the extraction of new variables, such as the time steps between measurements and time in milliseconds, as well as artefact removal, such as blinks.

The sound files had different lengths. To make sure this was not a problem in the analysis the time was set to 0 at the variable onset. The 200 ms before this were used to calculate the baseline. The baseline for each trial was calculated as the mean pupil size over these 200 ms, and was subtracted from the pupil size. By doing this, we could compare the relative changes in pupil size for each trial.

The statistics used to analyze the data collected in this experiment were Generalized Additive Mixed Models (GAMMs). GAMMs are essentially regression models, with the exception that they are capable of modelling non-linear regressions [53]. Whereas a typical linear regression assumes a linear relationship between the dependent variable and a predictor, a GAMM models the relationship “as a smooth function, which can, but does not need to be linear” [54]. This type of statistical analysis is particularly useful for datasets with dynamic and time-series data like ours. As mentioned by Wieling [53], it is often the case that such complex data is simplified during analysis, which leads to the possibility of missing interesting patterns that are present in the data. By using GAMMs, we do not have to simplify our data, thus leaving the door open to find these patterns. In order to fit GAMMs to our data, we used the packages *mgcv* [55] and *itsadug* [56] in R [52].

7. Results

The original dataset consisted of 1886 trials of data (41 subjects x 46 items). As is convention, trials that contained too many blinks were excluded from the analysis. In a similar manner to Vogelzang et al. [35], we set the threshold for removal at 25% blinks. In other words, trials in which more than 25% of the data constituted a blink, were removed. In the remaining trials we then removed the blinks, as well as 8 data points before and after the blink, corresponding to roughly 65 ms before and after each blink. Finally, we checked whether the remaining dataset for each trial was at least 75% of the size of the original (trial) data set. Of the original dataset, 241 (12.78%) trials had to be removed meaning that more than 87% of all trials were able to be kept for analysis. Generally, blinks are interpolated in order to avoid missing data, but because GAMMs can deal with missing data [54], we did not use interpolation here. In this study, the *mgcv* package [55] and its function *bam* were used in the R environment [52] to fit GAMMs to estimate the effects of salience on pupil size.

For each category, we plotted pupil size against time, resulting in the graphs in Fig. 1 below. The dotted lines represent pupil size for the variants we hypothesized to be less salient, whereas solid lines represent pupil size for the variants' more salient counterparts. Pupil size

appears to be larger for the salient variants in the categories *Loudness*, *Acoustic Prominence*, and *Grammatical Gender*.

Since the different categories comprised different types of variables, GAMMs were fitted to each of the six categories separately. In all six models, mean pupil diameter was the dependent variable. The GAMM models fitted to the data investigate the effect of salience on pupil size for our different categories. To carry out this investigation, the following model specification was used for all categories: $\text{Pupil} \sim \text{Condition} + \text{s}(\text{Time}, \text{by} = \text{Condition}) + \text{s}(\text{Time}, \text{Event}, \text{bs} = \text{'fs'}, \text{m} = 1) + \text{s}(\text{XGazePosRightEye}, \text{YGazePosRightEye})$. We can separate the model into the following chunks:

- $\text{Pupil} \sim \text{Condition}$: the formula reflecting the model specification, that is, pupil size depends on Condition.
- $\text{s}(\text{Time}, \text{by} = \text{Condition})$: indicates that for each of the levels of Condition (i.e. salient and non salient), a nonlinear regression line has to be estimated.
- $\text{s}(\text{Time}, \text{Event}, \text{bs} = \text{'fs'}, \text{m} = 1)$: the random smooth for Event. Event is an interaction of Trial and Subject, and as such represents each unique trial-participant combination. Because we also included the general smooth of time (see previous) this represents random adjustments for the general smooths.
- $\text{s}(\text{XGazePosRightEye}, \text{YGazePosRightEye})$: The non-linear interaction that accounts for the changes in pupil size resulting from gaze position.

As such, this model specification indicates that the dependent variable in our model, Pupil (pupil size), is modelled by allowing for the non-linear effect of Time (in ms) for both conditions (salient and non-salient). The non-linear random effect of Time and Event (each unique trial-participant combination is an event) is included to account for the order in which events were presented to the participants. This is necessary, because participants tend to become less attentive during the course of the experiment, but also because pupil size tends to increase over time when people are looking at a bright screen. Furthermore, the non-linear pattern between the X and Y gaze positions (the location on the screen that participants are looking at) is included to account for different measures based on the angle of incidence, which might affect registered pupil size.

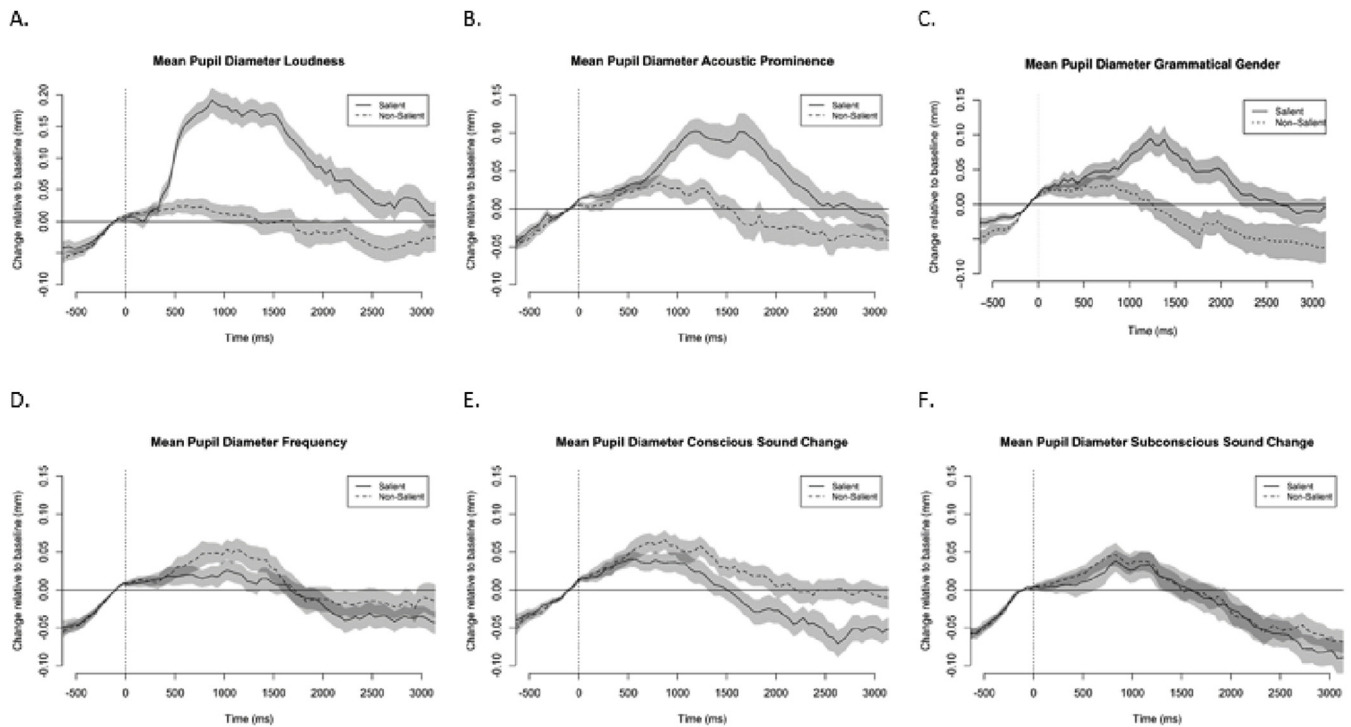
Finally, to control for autocorrelation in the residuals [53], we used the *itsadug* [56] function *acf_resid* and revised the model accordingly by filling the *bam* parameters *rho* and *AR.start*.

7.1. Loudness

The results of the model fit for the Loudness category can be found in Table 2. In rows 1 and 2, we see the estimates for the parametric coefficients. These show that the non-salient condition is 0.07 mm smaller than the baseline measured before stimulus onset. Furthermore, the salient condition is associated with larger pupil sizes (+0.09 mm) compared to the non-salient condition.

The remaining rows show the significance of the non-linear patterns associated with the condition over time, the non-linear random effect of time and event and the non-linear random effect of gaze position. The difference in pupil size over time, given in rows 3 and 4, is significant for both the salient ($F = 48.603, p < 0.001$) and non-salient ($F = 6.892, p < 0.001$) condition. The relatively high edfs (estimated degrees of freedom) for both rows show that the effect of time is highly non-linear for both conditions. Row 5 shows that the random effect of time and event is significant ($F = 207.217, p < 0.001$). The random effect of X and Y gaze positions in row 6 is also significant ($F = 1634.706, p < 0.001$), meaning that participants' gaze position indeed has an effect on measured pupil size.

In Fig. 2A, we see the expected change in pupil size over time for both conditions. In 2.B, we then see the difference in pupil size between those two conditions. We can see that there is a significant difference,



Mean pupil size over time for salient (solid) and non-salient (dashed) items per condition. The plots are centered around the start of the variables ($t = 0$). Change in pupil size is relative to the baseline (mean pupil size for 200ms preceding variable onset). (A) Mean pupil size for Loudness. (B) Mean pupil size for Acoustic Prominence. (C) Mean pupil size for Grammatical Gender. (D) Mean pupil size for Frequency. (E) Mean pupil size for Conscious Sound Change. (F) Mean pupil size for Subconscious Sound Change.

Fig. 1. Mean pupil size over time for salient (solid) and non-salient (dashed) items per condition. The plots are centered around the start of the variables ($t = 0$). Change in pupil size is relative to the baseline (mean pupil size for 200 ms preceding variable onset). (A) Mean pupil size for Loudness. (B) Mean pupil size for Acoustic Prominence. (C) Mean pupil size for Grammatical Gender. (D) Mean pupil size for Frequency. (E) Mean pupil size for Conscious Sound Change. (F) Mean pupil size for Subconscious Sound Change.

with bigger pupil sizes for the salient condition between approximately 400 ms–4000 ms.

In the interviews, all participants indicated that they perceive the salient condition as standing out because of its increased intensity. Furthermore, participants indicated that the stimuli were highly unexpected. Some mentioned that the change in intensity was somewhat startling at first. Participants agreed that the altered intensity resulted in an unnatural listening condition.

7.2. Acoustic prominence

The results of the model fit for the Acoustic Prominence category are shown in Table 3. In rows 1 and 2, we see the estimates for the parametric coefficients. Pupil size for the non-salient condition is 0.09 mm smaller than the baseline, measured before stimulus onset.

Table 2

Summary of the results of the GAMM model for the effect of Condition and Time for Loudness.

	Parametric coefficients:	Estimate	Std.error	t-value	Pr(> t)	
1	(Intercept)	-0.07005	0.01364	-5.134	2.84e-07	***
2	ConditionSalient	0.09348	0.01944	4.808	1.53e-06	***
	Smooth terms:	Edf	Ref.df	F	p-value	
3	s(Time):ConditionNonSalient	8.312	8.422	6.892	1.13e-09	***
4	s(Time):ConditionSalient	8.811	8.844	48.603	< 2e-16	***
5	s(Time, Event)	2085.048	2159.000	207.217	< 2e-16	***
6	s(XGazePositionRightEye, YGazePositionRightEye)	28.841	28.995	1634.706	< 2e-16	***

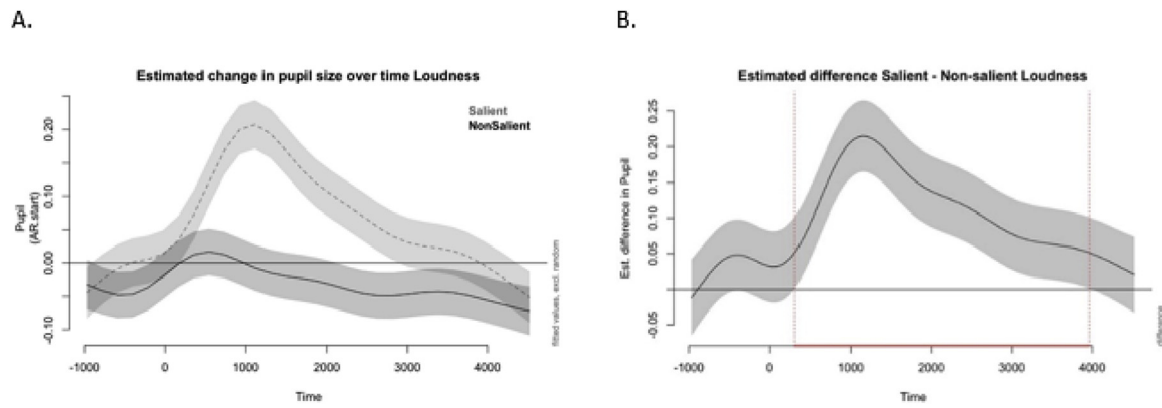
Asterisks indicates significance of: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The explained deviance of this model is 96%.

The positive estimate for the salient condition (row 2) indicates that overall, the salient condition is associated with larger pupil sizes (+0.05 mm) compared to the non-salient condition.

The remaining rows show the significance of the non-linear patterns associated with the condition over time, the non-linear random effect of Time and Event and the non-linear random effect of gaze position. The difference in pupil size over time, given in rows 3 and 4, is significant for both the salient ($F = 22.816, p < 0.001$) and non-salient ($F = 9.785, p < 0.001$) condition. The relatively high edfs for both conditions show that the effect of time is highly non-linear.

Row 5 shows that the random effect of time and event is significant ($F = 229.136, p < 0.001$). The random effect of X and Y gaze positions in row 6 is also significant ($F = 1989.894, p < 0.001$).

In Fig. 3A, we see the expected change in pupil size over time for both conditions, and in 3.B, the difference in pupil size between these.



The estimated effect over time and differences for Loudness. (A) The change in pupil size over time per condition (salient = solid, non-salient = dashed) as estimated by the model. (B) The difference between the pupil sizes for the different categories. Significant differences are marked in red.

Fig. 2. The estimated effect over time and differences for Loudness. (A) The Change in pupil size over time per condition (salient = solid, non-salient = dashed) as estimated by the model. (B) The difference between the pupil sizes for the different categories. Significant differences are marked in red.

We can note that there is a significant difference in the pupil sizes between salient and non-salient conditions, with bigger pupil sizes for the salient condition, between approximately 800 ms–2600 ms and 3400 ms–3900 ms.

During the interviews, all participants indicated that the condition that was hypothesized to be salient indeed stood out. The forms that were hypothesized to be salient were perceived as having flawed pronunciation. The stimuli that were pronounced with [p] instead of [t], were described as “strange”.

7.3. Grammatical gender

The results of the model fit for the Gender category can be found in Table 4. In rows 1 and 2, we see the estimates for the parametric coefficients. These show that the non-salient condition is 0.09 mm smaller than the base line measured before stimulus onset. The salient condition overall is associated with larger pupil sizes (+0.03 mm) compared to the non-salient condition.

The remaining rows show the significance of the non-linear patterns associated with the condition over time, the non-linear random effect of time and event, and the non-linear random effect of gaze position. We can see that the difference in pupil size over time, given in rows 3 and 4, is significant for both the salient ($F = 11.399, p < 0.001$) and non-salient ($F = 9.285, p < 0.001$) condition. The relatively high edfs for both rows show that the effect of time is highly non-linear for both conditions. Row 5 shows that the random effect of time and event is significant ($F = 278.799, p < 0.001$). The random effect of the X and Y gaze position in row 6 is also significant ($F = 1962.118, p < 0.001$).

In Fig. 4A, we see the expected change in pupil size over time for

both conditions, and in 4.B the difference in pupil size for both conditions. We can see that there is a significant variance, with bigger pupil sizes for the salient condition between approximately 1100 ms–2800 ms.

During the interviews, the violations of grammatical gender were reported to have stood out. Participants considered these variables to contain mistakes, and many stated that these were irritating, mainly because they felt that the use of such variations signals a low proficiency level of Dutch.

7.4. Frequency

The results of the model fit for the Frequency category showed no significant difference between pupil sizes for both categories.

During the interviews, more than half ($N = 24$) of the participants reported that the low frequent feature stood out. When asked why they found these variables more prominent, they specifically mentioned the terms’ low frequency. Other participants ($N = 15$) did not report differences in noticeability for the two conditions. The remaining participants ($N = 2$) found the more frequent words salient and stated that these were pronounced in a way that was not perceived as normal.

7.5. Conscious sound change

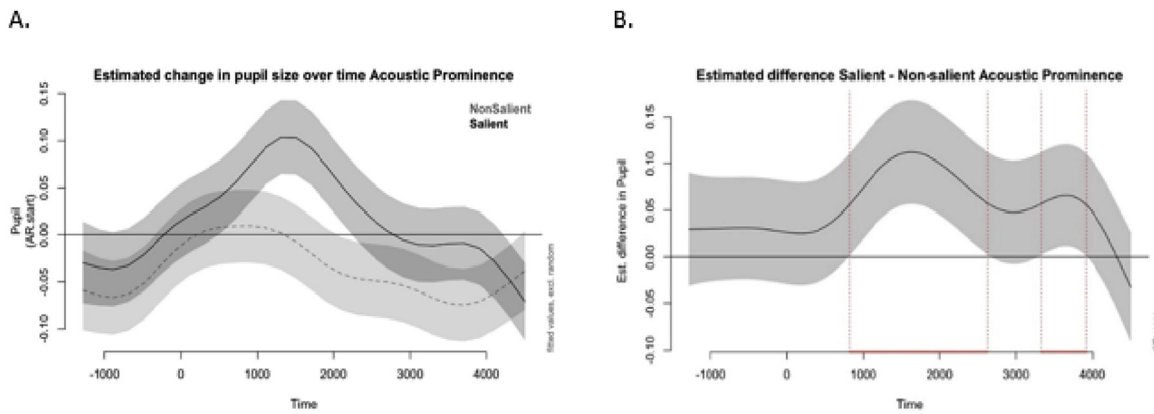
The results of the model fit for the Conscious Sound Change category show no significant difference between pupil sizes for both categories. Note that there is a trend of pupil size to be larger for the non-salient category (alveolar tap), yet this difference did not reach significance ($p = 0.09$).

Table 3

Summary of the results of the GAMM model for the effect of Condition and Time for Acoustic Prominence.

	Parametric coefficients:	Estimate	Std.error	t-value	Pr(> t)	
1	(Intercept)	-0.08743	0.01619	-5.400	6.68e-08	***
2	ConditionSalient	0.05368	0.02206	2.433	0.015	*
	Smooth terms:	Edf	Ref.df	F	p-value	
3	s(Time):ConditionNonSalient	8.098	8.240	9.785	4.15e-11	***
4	s(Time):ConditionSalient	8.496	8.578	22.816	< 2e-16	***
5	s(Time, Event)	2086.606	2167.000	229.136	< 2e-16	***
6	s(XGazePositionRightEye, YGazePositionRightEye)	28.105	28.807	1989.894	< 2e-16	***

Asterisks indicates significance of: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. The explained deviance of this model is 96%.



The estimated effect over time and differences for Acoustic Prominence. (A) The change in pupil size over time per condition (salient = solid, non-salient = dashed) as estimated by the model. (B) The difference between the pupil sizes for the different categories. Significant differences are marked in red.

Fig. 3. The estimated effect over time and difference for Acoustic Prominence. (A) The change in pupil size over time per condition (salient = solid, non-salient = dashed) as estimated by the model. (B) The difference between the pupil sizes for the different categories. Significant differences are marked in red.

During the interviews, a substantial portion of the participants reported the variant hypothesized to be non-salient, as actually standing out to them (N = 13). The majority (N = 23) did not favor one form over the other, although they did perceive the different variants and were able to point out the stereotype associated with the Gooise r. Only 5 participants reported the hypothesized salient variable as standing out. It turned out that most participants used the Gooise r themselves, or had a lot of exposure to it.

7.6. Subconscious Sound Change

The results of the model fit for the Subconscious Sound Change category showed no significant difference between pupil sizes for both categories. During the interviews, more than half of the participants (N = 21) were revealed to not have perceived a difference between the two variants. The vast majority of the participants (N = 31) did not report either of the two variables to be more noticeable than the other.

8. Discussion

The purpose of the present experiment was to find out whether salience could be measured in terms of pupil size, in order to gain a better understanding of the linguistic and cognitive correlates of salience. To do so, we tested whether the processing of properties reported to be salient in the literature would result in an increase in pupil diameter. Pupil size was measured while participants listened to sentences, half of which contained a trait that was hypothesized to be salient. Afterwards, mean pupil size for the different categories and conditions was compared.

Previous studies have shown that correlates of cognitive processing can be measured using neurophysiological measures (e.g. fMRI, EEG, pupillometry). Hence, we tested whether such measures could give us a quantifiable measure of salience, using pupil dilation as our variable. As stated by Beatty and Lucero-Wagoner [32], increases in cognitive load are reflected by changes in pupil size of up to 0.5 mms. We found such a result in three out of six linguistic categories. Although the differences identified are small, our results do suggest that there is an increase in cognitive load for the processing of Loudness, Acoustic Prominence and Grammatical Gender. The categories in which pupil dilation showed no significant response were Frequency, Conscious sound change and Sub-conscious sound change.

The results in the category Loudness showed that contrasts in stimulus intensity, elicited significantly larger pupil sizes. The link between loudness and auditory salience was previously pointed out by Liao et al. [12], and in this experiment we were able to replicate this type of pupillary responses for loudness using spoken samples (as opposed to Liao et al.’s study, where environmental sounds were used). Informants pointed out that this feature was salient due to its unexpectedness. This is in line with statements by, for example Zarcone et al. [9] and Jaeger and Weatherholtz [56], who argue that surprise is an important factor for a feature’s ‘salience’.

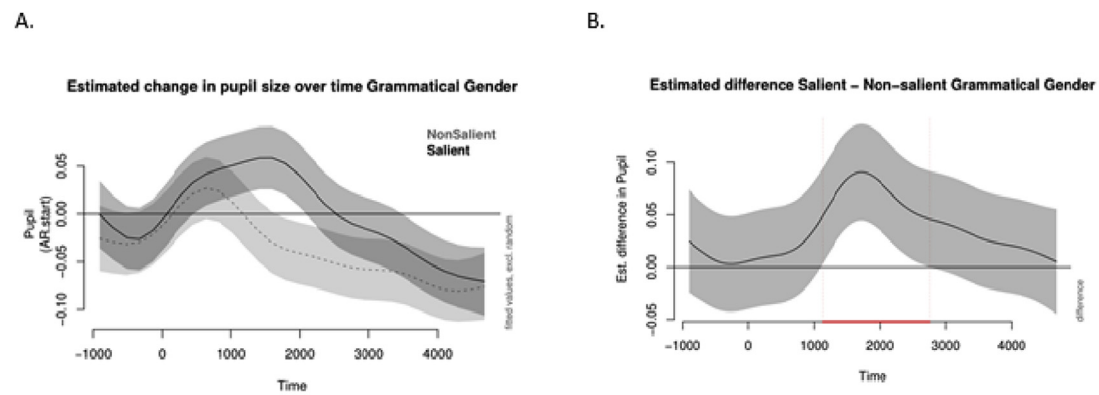
In the Acoustic Prominence category, the salient equivalents of the auditory sentence pairs consisted of a realization of a native phoneme in an unexpected position ([p] was put at the place where one would expect [t]). The salient condition indeed elicited significantly larger pupil sizes, confirming the belief that these variables are salient either due to their unexpectedness, or possibly as a result of error-recognition.

Similarly, the Grammatical Gender category revealed significantly

Table 4
Summary of the results of the GAMM model for the effect of Condition and Time for Grammatical Gender.

	Parametric coefficients:	Estimate	Std.error	t-value	Pr(> t)
1	(Intercept)	-0.08520	0.01260	-6.761	1.37e-11
2	ConditionSalient	0.03462	0.01752	1.977	0.0481
	Smooth terms:	Edf	Ref.df	F	p-value
3	s(Time):ConditionNonSalient	8.294	8.404	9.285	4.46e-12
4	s(Time):ConditionSalient	8.356	8.458	11.399	< 2e-16
5	s(Time, Event)	2372.991	2446.000	278.799	< 2e-16
6	s(XGazePositionRightEye, YGazePositionRightEye)	28.753	28.989	1962.118	< 2e-16

Asterisks indicates significance of: *p < 0.05, **p < 0.01, ***p < 0.001. The explained deviance of this model is 96%.



The estimated effect over time and differences for Grammatical Gender. (A) The change in pupil size over time per condition (salient = solid, non-salient = dashed) as estimated by the model. (B) The difference between the pupil sizes for the different categories. Significant differences are marked in red.

Fig. 4. The estimated effect over time and difference for Grammatical Gender. (A) The change in pupil size over time per condition (salient = solid, non-salient = dashed) as estimated by the model. (B) The difference between the pupil size for the different categories. Significant differences are marked in red.

larger pupil sizes for violations of grammatical gender. These syntactic violations have previously been called salient by Hanulíková et al. [19], who, in an ERP study, found an increased P600 effects, that is. a positive going wave observed after approximately 600 ms (hence the name P600) that is associated with syntactic violations. The authors reported a P600 response to gender agreement errors while processing native speech, but not while processing non-native speech. This suggests that, while violations made by a native speaker are perceived as salient, violations made by a non-native speaker, and in particular those who are known to make gender mistakes, are not. In other words, when listeners expect a certain accented speaker to be unable to correctly use gender marking, they also no longer repair their mistakes. It is unknown whether the same effect happens with regional accents. The increased pupil size found in our Grammatical Gender category is in line with the findings by Hanulíková et al. [19] and is thus likely to be an effect of the violations of syntactic expectations. This indicates that a part of what makes something salient, in the sense of an increase in pupil size, is error detection.

As discussed earlier in this paper, the literature on frequency and saliency is contradictory. Some argue that saliency is related to higher frequency [e.g. 7]. On the other hand, others propose low frequency is at the heart of saliency [e.g. 9]. When it comes to pupil dilation we find no significant difference for high frequency and low frequency conditions. A possible explanation for the absence of a frequency effect might be related to individual differences. Brysbaert, Mandera and Keuleers [23], for example, claimed the effect of frequency to be a highly personal one, based on personal language exposure. Thus, comparing results from multiple subjects might not be suitable for examining the effects of frequency on processing effort. This particular point requires more research.

One aim of this paper has been to consider the methodological benefits of neurophysiological measures for linguistic theory, particularly for sociolinguistics. In the Subconscious Sound Change category, in which we presented participants with voiced and voiceless representations of /v/, we found no significant differences in dilation degrees between our variants, as expected. Unexpectedly, the variant undergoing Conscious Sound Change, retroflex bunched r, showed no significant effect in terms of pupil dilation either. If anything, the data showed an opposite pattern to that which was hypothesized, with pupil size being slightly, though not significantly, larger for the 'non-salient' category (the alveolar tap) as compared to the category that was hypothesized to be salient (the retroflex bunched 'Gooise'/r/). It is possible that the group of participants was not large enough to find an

effect. Importantly, the qualitative interviews concerning the variation in /r/ indicate that the retroflex bunched variant carries social meaning. This social meaning difference between the two variants is not reflected in pupil dilation results. This result emphasizes the importance of work on understanding how social meaning and higher order indexicalities come about in language, as their importance for language change is clear. It is possible that this work cannot be done experimentally, but instead must be carried out qualitatively and longitudinally within speech communities, or communities of practice.

The question arising from these results concerns what pupillometry tells us here exactly. As discussed above, pupil dilation is inherently multi-causal. As such, we should carefully consider what could cause the pupil to dilate for these different categories, and whether this is a shared process or not. To answer this, we need to discuss what the three categories have in common. The interviews prove useful here, because they show that there might be two things underlying the pupil dilation. Firstly, participants found the variables in these three categories unexpected. Secondly, the variables were often seen as errors, and as such, the response might actually indicate that error recognition has taken place [57]. This brings forward a different point of interest, however. As discussed, saliency is associated with attentional processes. Indeed, salient variables seemed to be more noticeable across the categories (all except *Subconscious Sound Change*), but this was not reflected in the pupil dilation data. Thus, we might conclude that the capture of attention is not reflected by pupil dilation in our data.

In the introduction, we discussed how saliency aims to combine both structural (language-internal) factors with sociolinguistic and psychological (extra-linguistic) factors in a single explanatory concept [3]. Interestingly, the three categories that did show a significant result might be grouped as being language-internal, whereas the sound change categories for example can be grouped as being extra-linguistic. Possibly then, this would mean that pupil dilation is suited to testing these language-internal aspects, but not the extra-linguistic aspects of saliency.

Based on the results from this study, we have to conclude that the usefulness of pupillometry as a measure of saliency is questionable. Not all categories that were deemed salient by the participants elicited a (significantly) larger pupil size, suggesting that there is not one distinct cognitive process we might call 'saliency', but that there are in fact multiple distinct processes which may all result in the variable becoming more noticeable. While its usefulness is limited, we would not recommend sociolinguists discard the measure of pupillometry altogether. There are still possible applications of pupil dilation measures

for considering language change in progress. If a linguistic feature is innovative enough, its unexpectedness for the listener will be higher. In our study our informants happened to predominantly be users of the innovative r-variant themselves. The change to the innovative variant is not complete in all the age groups in the region, but the feature can still very much be said to be linked to younger speakers, cf. [58], who comprised the majority of our participant group. The fact that we measured the pupil responses in a young age group may have resulted in a lack of dilation effects due to the expectedness of the retroflex bunched variant. Furthermore, there are other neurophysiological measures that provide us with more detailed data, such as ERPs. These might shed more light on the processes at play in the processing of these salient variables.

As discussed in the introductory sections to this paper, the concept of linguistic salience faces multiple issues. A link should be made with the related notion of ‘markedness’, referring to the relative neutrality [59] or predictability [60] of a feature. Markedness, as introduced by Prague School linguist Trubetzkoy in 1931, is a broad notion that is used to distinguish between a marked and an unmarked form, the first of which is thought of as less neutral and/or less usual than the latter [59].

A detailed overview of markedness is given by Haspelmath [61], who observes the ambiguity of markedness and states that there are very few studies that make use of the concept in a way that encompasses all of its different definitions. Haspelmath recommends to stop using the concept altogether, since “simple everyday concepts should be expressed by simple everyday words” [61]. However, in spite of the ambiguity surrounding the concept of markedness, it is still central to many theories of language and phonology. Martins [62] concludes that “markedness seems to be the result of a conceptual mistake; it doesn’t really exist *per se*”. Perhaps, we as linguists should wonder to what extent the same is true for the concept of salience.

Our study, then, provides evidence that specific, but certainly not all, dimensions of salience can be measured using pupil size. We hypothesized that salient features would be more demanding to process, resulting in a larger processing load. Larger pupil sizes might serve as proof for increased processing load through the unexpectedness of the signal. We would also like to consider further whether the results are a reflection of error recognition. Especially for the Acoustic Prominence and Grammatical Gender categories, this would seem reasonable. In future work it would make sense to compare, and even combine, pupillometry measures with other measures, such as ERPs, to weed out the multi-causality effects that may occur in pupillometry. Such experiments could provide us with more detailed information about the responses elicited by our salient variables, and to the question of what pupil dilation is measuring exactly.

Appendix A. Transcript of The Stimuli per Category

Category		Salient	Non-salient
Grammatical Gender	List 1	Hij kent het naam hei 'kent het nam Hij staat voor het school hei stat vor het sxol Zij gaat naar de bedrijf zei xat nar də bə'dreif Zij zoeken de boek zei 'zūkən də buk	Zij werkt in de zaak zei 'werkt in də zak Zo eindigt de dag zo 'eindəxt də dāx Dat is het begin dat is het bə'xɪn Ik krijg de kans ik 'kreix də kəns
	List 2	Zij werkt in het zaak zei 'werkt in het zak Zo eindigt het dag zo 'eindəxt het dāx Dat is de begin dat is də bə'xɪn Ik krijg het kans ik 'kreix het kəns	Hij kent de naam hei 'kent də nam Hij staat voor de school hei stat vor də sxol Zij gaat naar het bedrijf zei xat nar het bə'dreif Zij zoeken het boek zei 'zūkən het buk
Acoustic Prominence	List 1	Ik heb veel gelp ik 'heb vel xelp Ik zie geen beelp ik zi xen belp Ze beginnen over één minuup ze bə'xɪnən 'ovər en mi'nyp Op de derde plaaps ɔp də 'dərde plaps	Ik heb benzine in mijn auto ik 'heb ben'zɪnə in meɪn 'auto Wij nemen de trein wei 'hemən də trein Zij maken een foto zei 'makən ən 'foto Hij wil een oneven aantal hei wɪl ən ən'evən 'antəl

Future work on salience should work towards exploring how innovations in language (unexpected by default) acquire new users. The answer to this may lie in the relationship between the feature and the social meaning attributed to it by innovators. Finally, we propose a move towards a theory of language variation and change in which the processes that occur in our cognitive systems are included.

9. Conclusion

In this paper we have reviewed various operationalizations of salience in which salience is either understood as standing out or being noticeable, or on the contrary as being most obvious, or logical in a given situation. According to the previous literature salience may be linked to surprisal, and this is in line with the results from our experimental, as well as qualitative, analysis.

In our eye-tracking experiment pupil size was found to significantly increase when participants were presented with salient traits in three of the six categories under survey in this experiment: *Acoustic Prominence*, *Loudness* and *Grammatical Gender*. Although other categories showed similar trends, these were not found to be significant. Based on these results, we are able to conclude that salience can, in some cases, be measured in pupil size. This is likely to be a result of the level of surprisal that imposes an increased load on the cognitive system. Zarcone et al. [9] have previously pointed towards this possible relationship between salience and surprisal. The relationship between salience and surprisal is however still unclear and deserves more attention in future research.

Although a univocal definition of salience is still lacking, the results in this experiment have helped us come closer to disentangling the concept of salience by showing that specific salient traits show a physiological effect measurable in pupil sizes. This trait is linked to surprisal and unexpectedness. The categories that were chosen in this experiment all fitted criteria for salience presented by various scholars in the field. Our lack of significant results for some of these categories suggests that existing definitions of salience need reevaluation. Alternatively, we as linguists should consider abandoning the term altogether, as the definitions in use at the moment are conflicting and far from clear-cut.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

	List 2	Ik heb benzine in mijn aupto ik 'heb ben'zine in mein 'aupto Wij nemen de prein wei 'nemən də prein Zij maken een fopo zei 'makən ən 'fopo Hij wil een oneven aanpal hei wil ən ɔn'evən 'anpal	Ik heb veel geld ik 'heb vel xelt Ik zie geen beeld ik zi xen belt Ze beginnen over één minuut ze bə'xɪnən 'ovər en mi'nyt Op de derde plaats ɔp də 'dɛrdə plɑts Zij stelt een vraag zei stelt ən frax Daar ligt een leeg vel dar lixt ən lex fel Iedereen kent het gevoel 'idərən 'kent et xə'ful In die omgeving in di ɔm'xevɪŋ Hij heeft een vrouw hei 'heft ən frau Dat is een val dat is ən val Alles heeft een gevolg 'aləs 'heft ən xə'vɔlx Dat in geen geval dat in xen xə'val
Subconscious Sound Change	List 1	Hij heeft een vrouw hei 'heft ən vrau Dat is een val dat is ən val Alles heeft een gevolg 'aləs 'heft ən xə'vɔlx Dat in geen geval dat in xen xə'val	Zij stelt een vraag zei stelt ən frax Daar ligt een leeg vel dar lixt ən lex vel Iedereen kent het gevoel 'idərən 'kent et xə'vul In die omgeving in di ɔm'xevɪŋ Hij heeft een vrouw hei 'heft ən frau Dat is een val dat is ən val Alles heeft een gevolg 'aləs 'heft ən xə'vɔlx Dat in geen geval dat in xen xə'val
	List 2	Zij stelt een vraag zei stelt ən vrax Daar ligt een leeg vel dar lixt ən lex vel Iedereen kent het gevoel 'idərən 'kent et xə'vul In die omgeving in di ɔm'xevɪŋ	Dat is zijn moeder dat is zein 'mudər Op dat late uur ɔp dat 'latə yr Wij zijn al ouder wei zein al 'audər Ons favoriete nummer ɔns favo'rite 'nymər Op die manier ɔp di ma'nir De volgende keer də 'vɔlxəndə kɛr Die twee zijn een paar di twe zein ən pər Een open deur ən 'ɔpən dɔr
Conscious Sound Change	List 1	Op die manier ɔp di ma'nir De volgende keer də 'vɔlxəndə kɛr Die twee zijn een paar di twe zein ən pər Een open deur ən 'ɔpən dɔr	Ons favoriete nummer ɔns favo'rite 'nymər Op die manier ɔp di ma'nir De volgende keer də 'vɔlxəndə kɛr Die twee zijn een paar di twe zein ən pər Een open deur ən 'ɔpən dɔr
	List 2	Dat is zijn moeder dat is zein 'mudər Op dat late uur ɔp dat 'latə yr Wij zijn al ouder wei zein al 'audər Ons favoriete nummer ɔns favo'rite 'nymər	Het eerste deel het 'erstə del Een kleine jongen ən kleinə 'jɔŋən Dat lelijke ding dat 'leləkə dɪŋ Zij heeft veel zin zei 'heft vel zin
Loudness	List 1	De komende week də 'koməndə wɛk Een goede reden ən xudə 'redən Ons geweldige idee ɔns xə'wɛldəxə 'ide Hij is het probleem hei is het 'prɔblɛm	Het eerste deel het 'erstə del Een kleine jongen ən kleinə 'jɔŋən Dat lelijke ding dat 'leləkə dɪŋ Zij heeft veel zin zei 'heft vel zin
	List 2	Het eerste deel het 'erstə del Een kleine jongen ən kleinə 'jɔŋən Dat lelijke ding dat 'leləkə dɪŋ Zij heeft veel zin zei 'heft vel zin	De komende week də 'koməndə wɛk Een goede reden ən xudə 'redən Ons geweldige idee ɔns xə'wɛldəxə 'ide Hij is het probleem hei is het 'prɔblɛm
Frequency	List 1	Een lastig doel (1) ən 'lastəx dɔl Een hoop druk (1) ən hɔp drɪk Hij kent de regel (1) hei kent də 'rɛxəl Hij kiest voor het belang (1) hei kist vor het bə'laŋ	Een groot hol (1) ən xrot hɔl Ik zie een huis (1) ik zi ən hɔɛys Zijn dikke zoon zein dikə zɔn Dat doen wij met gemak dat dun wei met xə'mək
	List 2	Een grote rol ən xrotə rɔl Ik zie een huis ik zi ən hɔɛys Zijn dikke zoon zein dikə zɔn Dat doen wij met gemak dat dun wei met xə'mək	Een lastig duel ən 'lastəx dy'wɛl Een hoop drek ən hɔp drɛk Hij kent het zegel hei kent het 'zɛxəl Hij kiest voor het behang hei kist vor et bə'haŋ

References

- [1] M. Meyerhoff, *Introducing Sociolinguistics*, 2, Routledge, London, 2011.
- [2] R. Hickey, Saliency, stigma and standard, in: L. Wright (Ed.), *The Development of Standard English 1300–1800*, Cambridge University Press, Cambridge, 2000, pp. 57–72, <https://doi.org/10.1017/CBO9780511551758.005>.
- [3] P. Kerswill, A. Williams, "Saliency" as an explanatory factor in language change: evidence from dialect levelling in urban England, *Language Change: the Interplay of Internal, External and Extra-linguistic Factors*, De Gruyter Mouton, Berlin, 2002, <http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=nlebk&AN=558656&site=ehost-live&scope=site>.
- [4] P. Rącz, *Saliency in Sociolinguistics: a Quantitative Approach*, De Gruyter Mouton, Berlin, 2013 <https://www-degruyter-com.proxy-ub.rug.nl/viewbooktoc/product/203926>.
- [5] P. Trudgill, *Dialects in Contact*, B. Blackwell, Oxford, 1986 UK ; New York, NY, USA.
- [6] V. Schirmunski, *Sprachgeschichte und Siedlungsmundarten*, *Germanisch-Romanische Monatsschrift* 18 (1930) 171–188 113–122.
- [7] R. Giora, *On Our Mind Saliency, Context, and Figurative Language*, Oxford University Press, 2003, <https://doi.org/10.1093/acprof:oso/9780195136166.001.0001>.
- [8] K. Franco, *Concept Features and Lexical Diversity: A Dialectological Case Study on the Relationship between Meaning and Variation*, KU Leuven, 2017, https://lirias.kuleuven.be/bitstream/123456789/587898/1/doctoraat_karlien.pdf.
- [9] A. Zarcone, M. van Schijndel, J. Vogels, V. Demberg, Saliency and attention in surprisal-based accounts of language processing, *Front. Psychol.* 7 (2016), <https://doi.org/10.3389/fpsyg.2016.00844>.
- [10] N.C. Ellis, Saliency, language complexity, and complex adaptive systems, *Stud. Sec. Lang. Acquis.* 38 (2016) 341–351, <https://doi.org/10.1017/S027226311600005X>.
- [11] A. Blumenthal-Dramé, A. Hanulíková, B. Kortmann, Editorial: perceptual linguistic saliency: modeling causes and consequences, *Front. Psychol.* 8 (2017), <https://doi.org/10.3389/fpsyg.2017.00411>.
- [12] H.-I. Liao, S. Kidani, M. Yoneya, M. Kashino, S. Furukawa, Correspondences among pupillary dilation response, subjective saliency of sounds, and loudness, *Psychon. Bull. Rev.* 23 (2016) 412–425, <https://doi.org/10.3758/s13423-015-0898-0>.
- [13] S.D. Goldinger, M.H. Pappas, Pupil dilation reflects the creation and retrieval of memories, *Curr. Dir. Psychol. Sci.* 21 (2012) 90–95, <https://doi.org/10.1177/0963721412436811>.
- [14] J. Hyönä, J. Tommola, A.-M. Alaja, Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks, *Q. J. Exp. Psychol. Sect. A* 48 (1995) 598–612, <https://doi.org/10.1080/14640749508401407>.
- [15] T. Koelewijn, H. de Kluijver, B.G. Shinn-Cunningham, A.A. Zekveld, S.E. Kramer, The pupil response reveals increased listening effort when it is difficult to focus attention, *Hear. Res.* 323 (2015) 81–90, <https://doi.org/10.1016/j.heares.2015.02>.

- 004.
- [16] F. Paas, J.E. Tuovinen, H. Tabbers, P.W.M. Van Gerven, Cognitive load measurement as a means to advance cognitive load theory, *Educ. Psychol.* 38 (2003) 63–71, https://doi.org/10.1207/S15326985EP3801_8.
- [17] U. Weinreich, W. Labov, M.I. Herzog, Empirical foundations for a theory of language change, *Directions for Historical Linguistics*, Univeristy of Texas Press, Austin, Texas, 1968, pp. 95–195 <https://www.staff.ncl.ac.uk/joel.wallenberg/ContextsJoelGeoff/WeinreichLabovHerzog1968.pdf>.
- [18] C. Llamas, D. Watt, A.E. MacFarlane, Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone, *Front. Psychol.* 7 (2016), <https://doi.org/10.3389/fpsyg.2016.01163>.
- [19] A. Hanulíková, P.M. van Alphen, M.M. van Goch, A. Weber, When one person's mista is another's standard usage: the effect of foreign accent on syntactic processing, *Al. Cognit. Neurosci.* 24 (2012) 878–887.
- [20] H. Loerts, Uncommon Gender. Eyes and Brains, Native and Second Language Learners, and Grammatical Gender, Grodil, Groningen, 2013.
- [21] A. Cieśllicka, Literal salience in on-line processing of idiomatic expressions by second language learners, *Sec. Lang. Res.* 22 (2006) 115–144, <https://doi.org/10.1191/0267658306sr263oa>.
- [22] C. Chapman, Perceptual salience and analogical change: evidence from vowel lengthening in modern Swiss German dialects, *J. Ling.* 1 (1995) 1–13.
- [23] M. Brysbaert, P. Mandera, E. Keuleers, The word frequency effect in word processing: an updated review, current directions in psychological science, 27 (2018) 45–50, <https://doi.org/10.1177/0963721417727521>.
- [24] W. Labov, *Sociolinguistic Patterns*, 11. Print, Univ. of Pennsylvania Press, Philadelphia, 1991.
- [25] R. Bezooijen, R. van den Berg, R. De Goosie, Wie ziet er wat in en waarom? Taaivariatie En Groepsidentiteit 17 (2004), <http://repository.uibn.ru.nl/bitstream/handle/2066/40866/40866.pdf?sequence=1>.
- [26] K. Sebregts, The Sociophonetics and Phonology of Dutch R, Utrecht University, 2015, <https://dspace.library.uu.nl/handle/1874/306415>.
- [27] P. Rácz, Operationalising salience: definite article reduction in the North of England, *Engl. Lang. Ling.* 16 (2012) 57–79, <https://doi.org/10.1017/S1360674311000281>.
- [28] A.-F. Pinget, The Actuation of Sound Change, Utrecht University, 2015, <https://dspace.library.uu.nl/handle/1874/318102> accessed February 22, 2017.
- [29] B. Johansson, C. Balkenius, A computational model of pupil dilation, *Connect. Sci.* (2017) 1–15, <https://doi.org/10.1080/09540091.2016.1271401>.
- [30] H. van Rijn, J.R. Dalenberg, J.P. Borst, S.A. Sprenger, Pupil dilation Co-varies with memory strength of individual traces in a delayed response paired-associate task, *PLoS One* 7 (2012) e51134, , <https://doi.org/10.1371/journal.pone.0051134>.
- [31] C.J. Ellis, The pupillary light reflex in normal subjects, *Br. J. Ophthalmol.* 65 (1981) 754–759, <https://doi.org/10.1136/bjo.65.11.754>.
- [32] J. Beatty, B. Lucero-Wagoner, The pupillary system, *Handbook of Psychophysiology*, Cambridge University Press, Cambridge, 2000, pp. 142–162 http://apps.usd.edu/coglab/schieber/docs/Beatty_Pupillary_System_.pdf.
- [33] E.H. Hess, J.M. Polt, Pupil size in relation to mental activity during simple problem-solving, *Science* 143 (1964) 1190–1192, <https://doi.org/10.1126/science.143.3611.1190>.
- [34] D. Kahneman, J. Beatty, Pupil diameter and load on memory, *Science* 154 (1966) 1583–1585, <https://doi.org/10.1126/science.154.3756.1583>.
- [35] M. Vogelzang, P. Hendriks, H. van Rijn, Pupillary responses reflect ambiguity resolution in pronoun processing, *Language, Cognit. Neurosci.* 31 (2016) 876–885, <https://doi.org/10.1080/23273798.2016.1155718>.
- [36] S. Mathôt, J. Grainger, K. Strijkers, Pupillary responses to words that convey a sense of brightness or darkness, *Psychol. Sci.* (2017) 095679761770269, , <https://doi.org/10.1177/0956797617702699>.
- [37] M.M. Bradley, L. Miccoli, M.A. Escrig, P.J. Lang, The pupil as a measure of emotional arousal and autonomic activation, *Psychophysiology* 45 (2008) 602–607, <https://doi.org/10.1111/j.1469-8986.2008.00654.x>.
- [38] W. Einhäuser, C. Koch, O. Carter, Pupil dilation betrays the timing of decisions, *Front. Hum. Neurosci.* (2010), <https://doi.org/10.3389/fnhum.2010.00018>.
- [39] L. Kuchinke, M. Vo, M. Hofmann, A. Jacobs, Pupillary responses during lexical decisions vary with word frequency but not emotional valence, *Int. J. Psychophysiol.* 65 (2007) 132–140, <https://doi.org/10.1016/j.ijpsycho.2007.04.004>.
- [40] R. Brünken, T. Seufert, F. Paas, Measuring cognitive load, in: J.L. Plass, R. Moreno, R. Brunken (Eds.), *Cognitive Load Theory*, Cambridge University Press, Cambridge, 2010, pp. 181–202, , <https://doi.org/10.1017/CBO9780511844744.011>.
- [41] J.J.A. van Berkum, *The Psycholinguistics of Grammatical Gender: Studies in Language Comprehension and Production*, Max Planck Institute for psycholinguistics, 1996.
- [42] H. Loerts, M. Wieling, M.S. Schmid, Neuter is not Common in Dutch: Eye Movements Reveal Asymmetrical Gender Processing, *J. Psycholinguist. Res.* 42 (2013) 551–570, <https://doi.org/10.1007/s10936-012-9234-2>.
- [43] J. Rommers, Online effecten van woordfrequentie en zinscontext: een ERP-studie, BA thesis Rijksuniversiteit Groningen, 2007.
- [44] C. Tiberius, T. Schoonheim, *A Frequency Dictionary of Dutch: Core Vocabulary for Learners*, Routledge, Taylor & Francis Group, London ; New York, 2014.
- [45] N. Oostdijk, W. Goedertier, *Corpus Gesproken Nederlands*, Instituut van de Nederlandse Taal, 2003.
- [46] G. Bouma, *Twitter Ngrams*, Univeristy of Groningen, 2014.
- [47] H. van de Velde, van Hout, The pronunciation of (r) in standard Dutch, *Ling. Neth.* 16 (1999) 177–188, <https://doi.org/10.1075/avt.16.16van>.
- [48] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer, (2017) www.praat.org.
- [49] Psychology Software Tools, Inc, E-Prime 2.0, Psychology Software Tools, Inc, 2012, www.pstnet.com.
- [50] Psychology Software Tools, Inc, E-Prime Extensions for Tobii, Psychology Software Tools, Inc, Pittsburgh, 2009.
- [51] J. Elvesjö, M. Skogö, H. Eskilsson, Tobii Eye Tracker, Tobii Technology AB, n.d.
- [52] R. R Core Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017 <https://www.R-project.org>.
- [53] M. Wieling, Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English, *J. Phonetics* 70 (2018) 86–116, <https://doi.org/10.1016/j.wocn.2018.03.002>.
- [54] J. van Rij, P. Hendriks, H. van Rijn, R.H. Baayen, S.N. Wood, Analyzing the time course of pupillometric data, *Trends Hear.* 23 (2019) 1–22, <https://doi.org/10.1177/2331216519832483>.
- [55] S.N. Wood, Package “Mgcv”, (2018) <https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>.
- [56] J. van Rij, M. Wieling, H. Baayen, H. Van Rijn, Package “Itsadug”, (2017) <https://cran.r-project.org/web/packages/itsadug/itsadug.pdf>.
- [57] P.R. Murphy, M.L. van Moort, S. Nieuwenhuis, The pupillary orienting response predicts adaptive behavioral adjustment after errors, *PLoS One* 11 (2016) e0151763, , <https://doi.org/10.1371/journal.pone.0151763>.
- [58] N. Haug Hilton, E.M. Smidt, L. Rosseel, A. Nota, M. Coler, Northern/R-/Variation in Dutch and the Social Meaning of the “Goosie R, Presented at: Dei fan de Fryske taalkunde/Dag van de Friese taalkunde, Fryske Akademy, Leeuwarden, 2015.
- [59] R.L. Trask, P. Stockwell, *Language and Linguistics: the Key Concepts*, second ed., Routledge, Abingdon [England], New York, 2007.
- [60] E. Hume, Deconstructing Markedness: A Predictability-Based Approach, (2004) <file:///C:/Users/MWP-WKS035361/Downloads/948-921-1-PB.pdf>.
- [61] M. Haspelmath, Against markedness (and what to replace it with), *J. Ling.* 42 (2006) 25, <https://doi.org/10.1017/S0022226705003683>.
- [62] P.T. Martins, There is no place for markedness in biologically-informed phonology, in: B.D. Samuels (Ed.), *Linguist. AktuellLinguistics Today*, John Benjamins Publishing Company, Amsterdam, 2017, pp. 219–232, , <https://doi.org/10.1075/la.241.09mar>.