How hard can it be to ignore the pan in *pan*da? Effort of lexical competition as measured in pupil dilation

Anita Wagner, Paolo Toffanin & Deniz Başkent

University of Groningen, University Medical Center Groningen, Dep. Otorhinolaryngology, Groningen, The Netherlands
a.wagner@umcg.nl

ABSTRACT

When speech comprehension becomes effortful, e.g. in adverse conditions, it is necessary to understand which processes are the sources of the effort. By means of pupillometry, this paper investigates whether the automatic process of lexical competition increases effort, and whether mismatching cues, degradation of the signal, and listener' experience with degraded signals additionally affect the effort, as reflected by increased pupil dilation.

Listeners' pupil dilation was measured during the disambiguation of embedded words when presented with cues that either served or hampered disambiguation. Furthermore, listeners were presented with natural or degraded speech. The effect of experience was measured by testing also long-term users of cochlear implants.

Lexical competition increased the effort of processing natural speech and even more so for stimuli with mismatching cues. Mere listening to degraded speech increased the effort more than lexical competition. Experience with degraded speech reduced the effort stemming from listening to un-natural speech.

Keywords:
Speech Comprehension, Phonetic Psycholinguistics, Lexical Competition, Degraded Speech

1. INTRODUCTION

Various mental and lexical processes are necessary for successful speech comprehension. Nevertheless, speech perception is commonly seen as effortless, because these mental processes operate automatically, and are facilitated by listeners' perceptual specialization to their native language [4], which guides their attention to phonetic cues [2]. In addition, the phonetically rich nature of the speech signal provides listeners with fine phonetic details, which govern automatic processes, such as lexical competition [5]. When speech perception is hampered, either by degraded signals, by mismatching cues, or by hearing impairment, speech comprehension becomes effortful. To find solutions to reduce this effort, it is important to understand how degrading factors contribute to the effort, and to capture it in an objective measure.

The measurement of pupil dilation has been used to study effort involved in solving various cognitive tasks [2, 1]. How emotional, cognitive and individual factors interact in reflecting mental processes leading to pupil dilations is not entirely known, but pupillometry has been shown to reflect the cognitive effort involved in listening to speech. An increase in pupil dilation can be observed when listeners are presented with degraded speech relative to highly intelligible speech [8]. Pupil dilation is an online measure of processing effort, even though it is a slow response, starting at about 900 ms after the onset of the effortful event [1].

Lexical competition is a fundamental process of speech comprehension that enables listeners to map spoken words to their mental representations which best match the heard speech signal. The selection of appropriate acoustic cues has been shown to govern lexical competition, and attention to irrelevant cues, e.g. non-native cues, increases lexical competition [7]. In addition, also fine phonetic details govern lexical competition very early upon hearing words [5].

A situation in which fine phonetic details versus phonetic cues are playing against each other is the case of users of cochlear implants (CIs). These listeners are otherwise deaf and have to rely on the spectrotemporally-degraded speech transmitted by their device for speech communication. The CI allows them to perceive speech because sufficient relevant cues are present in an otherwise strongly degraded signal [6], and CI users are likely to learn over time to make use of the transmitted cues. Worldwide more than 300000 listeners regained speech comprehension thanks to CIs, but for many of them speech processing is a tiring task. This raises the question of which automatic processes are the source of effort that remains unnoticed when speech is perceived under most optimal conditions.

Lexical competition taking place during the disambiguation of lexical embeddings (e.g., *pain* in *pain*ting) was studied by Salverda et al [5]. In their eye fixations study these authors showed that disambiguation is guided by fine phonetic details that listeners interfere from the prosody of a sentence. The crucial cues for early disambiguation were the durational differences between syllables in monosyllabic versus bisyllabic words. This study provides a starting point to disentangle the contribution of detailed cues in degraded speech, since duration can be well preserved in degraded signals, at least in the signal transmitted via CIs.

To investigate how lexical competition with natural and degraded speech contributes to processing effort we adapt the paradigm by Salverda et al. [5], and record pupil dilation alongside of gaze fixations, which indicate the time course of lexical competition. This allows us to address the following three questions. (1) Is lexical competition reflected in pupil dilation, and is pupil dilation also sensitive to increased lexical competition in the presence of mismatching cues? (2) Can signals with degraded acoustic form, in which however the crucial durational cues are preserved, lead to comparable pupil dilation patterns? (3) Does experience with degraded speech change the pupil response reflecting lexical competition and processing of degraded speech? The last point also implicates an answer to the question of whether CI users can process speech by means of the same mechanisms as normal hearing listeners do.

2. EXPERIMENT

2.1. Participants

60 native-Dutch speaker volunteers, aged between 19 and 27 (mean = 24 yrs.), participated in this study. None of them reported any known hearing or learning difficulties. Their hearing thresholds were normal, i.e. below 20 dB HL at the audiometric frequencies between 500 and 8000 kHz. Half of these volunteers participated in a task with natural speech (NS), and the other half with degraded speech (DS). In addition, 24 CI users, aged between 34 and 80 (mean = 62 yrs.) were also tested with the natural stimuli (CI). These participants were using their CI for at least twelve months.

2.2. Materials

The materials of Salverda [5] were rerecorded by a female native speaker of Dutch, and in analogy to [5] acoustically manipulated to create two conditions: target-matching condition, in which the durational cues matched the target, and competitor-matching condition, in which the durational cues matched the competitor. An example of the sentences and the manipulation is shown in Table 1. The stimuli set consisted of 26 sentence pairs with a target (e.g. pinda) and a competitor (pin) recorded in the same phonetic context, regarding the gross phonetic sequences, including intonation and stress patterns. Pictures of the target and competitor, together with two unrelated distractor pictures, were presented on a display (Figure 2). Next to the experimental items also 40 filler sentences were recorded and manipulated analogously. The fillers were displayed with either no competitor at all, or with a monosyllabic target and bisyllabic competitor, or with a competitor with final embedding (e.g. *pain* in cham*pagne*).

Table 1: An example of the recorded sentences, and the splicing manipulation.

Sentence 1	Ik wilde de PINDA opeten	
Sentence 1	Ik wilde de PINDA opeten	
Sentence 2	Ik wilde de PIN daarachter steken	
Target cues	Ik wilde de PIN	DA opeten
Competitor cues	Ik wilde de PIN	DA opeten

To create degraded stimuli the natural sentences were vocoded using an acoustic simulation of CIs. The stimuli were first bandlimited to 100 to 8000 Hz, and were subsequently bandpass filtered into 8 channels. The slow-varying amplitude envelopes were extracted in each channel via lowpass filtering, and these envelopes then modulated carrier sinusoid tones. The processed stimuli were the summed signals from the output of all channels. This manipulation led to stimuli that preserved the durational cues in an otherwise unnatural and spectrotemporally degraded acoustic form. Figure 1 display excerpts from spectrograms of the natural and degraded stimuli.

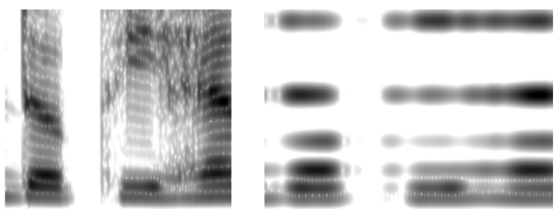

Figure 1: Excerpts from spectrogram of the natural stimuli (left), and the degraded stimuli (right).

2.3. Procedure

Before data collection, the participants were familiarized with the pictures used, and the words that referred to the pictures. They were then seated in a comfortable chair facing the monitor, and the eye-tracker was mounted on their head and calibrated. The task was to listen to the stimuli and to click on the object mentioned in the sentence. The pupil size was recorded together with gaze fixations by means of en Eyelink 500 eye-tracker. All listeners were presented with 4 practice trials, and 66 experimental trials, containing 26 trials with both conditions, but only one condition per experimental item.

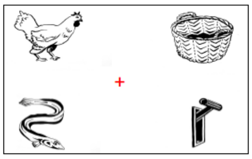

Figure 2: Example of the display presented to the participants, with hen as competitor for handle.

2.4. Analysis

Trials in which participants clicked on the wrong picture were excluded from the analysis. Trials with blinks longer than 300 ms were also excluded. Shorter eye blinks were corrected by means of linear interpolation. To investigate the effect of lexical competition with matching and mismatching cues the pupil dilation data per participant were baseline corrected to the 200 ms preceding the presentation of the experimental item (Baseline1). To investigate whether the effect of listening per se caused pupil dilation, the interval of 200 ms preceding the presentation with the very first sentence was chosen as a second baseline (Baseline 2). The baseline corrected data were normalized to correct for individual differences in pupil size, according to Equation 1for **E**vent **R**elated **P**upil **D**ilation (ERPD).

$$(1)\quad \%\, ERPD = (observation - baseline)/baseline * 100$$

2.5. Results

The pupil dilation data were analysed in their time course, as polynomial curves of 3rd order. The terms describing the curves are: intercept, the slope, and a coefficient for the curvature. These time curves were analysed by means of multilevel nonlinear regression models. The statistical models contained next to the terms describing the curve per participant also an interaction with these three terms per participant and random effects on these terms per participant.

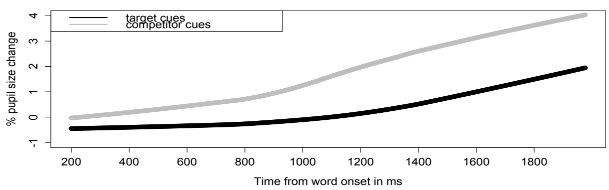

Figure 3: Pupil dilation time curves for experiment NS.

For display purposes in Figures 3,4, and 5, the time curves of dilation for the two conditions (target or competitor matching cues) were subtracted from the curves for the fillers, to reflect the effect of lexical competition, which was minimized in the fillers.

The results for the experiment with natural speech (NS) are displayed in Figure 3. The curves show the increase in pupil size over time as a function of lexical competition. The statistical analysis revealed that the target-matching cue curves differed from the competitor-matching cue curves on all terms describing the curves (χ^2 (1)= 35.89, p < .001). The curves for both conditions also differed significantly from the filler items in terms of slope (χ^2 (1)= 5.99, p < .001), curvature (χ^2 (1)= 8.65, p < .001), and area under the curve (χ^2 (1)= 16.53, p < .001). This implies that pupil dilation was sensitive to capture the effect of lexical competition, and that dilation was significantly bigger when the cues were mis-matching the target.

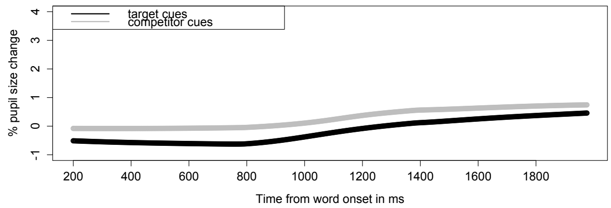

Figure 4: Pupil dilation time curves for experiment DS.

Figure 4 displays the pupil dilation time course for the experiment with degraded speech (DS). These curves differed from each other only in terms of their intercept (χ^2 (1)= 18.3, p < .001), and both differed from the filler items only in the curvature of the function (χ^2 (1)= 5.84, p < .001). This suggests that pupil dilation here did not capture effects of cue manipulation, and that the effect of lexical competition was only marginal.

Figure 5 displays the results for the CI participants. The functions for both conditions differed significantly form each only in the term slope (χ^2 (1)= 18.06, p < .001), and from the filler curves in terms of intercept (χ^2 (1)= 11.22, p < .001), and slope (χ^2 (1)= 4.25, p < .001). CI participants show an effect of lexical competition, but no effect of mismatching cues.

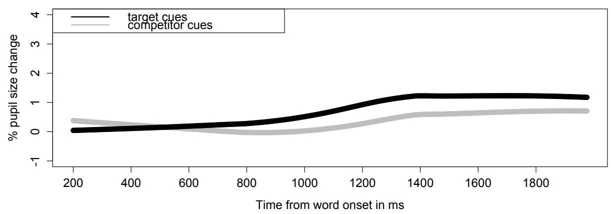

Figure 5: Pupil dilation time curves for experiment CI.

A final model compared the dilation time curves across participant groups and conditions. The NS results significantly differed from the DS results in all terms describing the curves. The NS results differed from the CI results in terms of slope of the curve and area under the curve. The main differences between NS and CI were thus the smaller increase in pupil dilation as a function of lexical competition, and no effect of mismatching cues for CI. CI results differed significantly from the DS results in terms of slope of the curves and intercept, showing an even smaller increase in dilation in DS than CI.

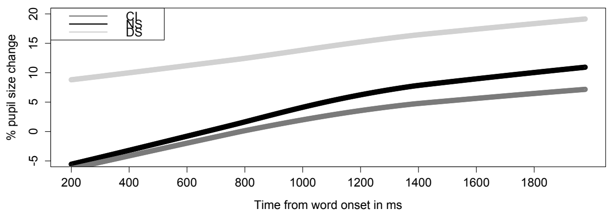

Figure 6: Pupil dilation time curves for all participants with baseline 2, reflecting the effect of listening per se.

Figure 6 displays the curves for all participants for Baseline 2. No effect of cue manipulation was found, hence only one curve is displayed per participant group. Normal hearing listeners did not differ from CI listeners, but participants in DS showed significantly greater pupil dilation as a response to the processing of the sentences per se.

3. GENERAL DISCUSSION

The results show that pupillometry can capture the effort related to the automatic processing of lexical competition, and also reflect differences in the processing of matching versus mismatching cues. Pupil dilation also captured the effort related to processing degraded speech (Baseline 2), which supports the results by Zekveld et al. [8]. For the NS and CI listeners there was no additional increase in dilation due to the processing of the sentences per se. For the DS listeners, the strong response to degraded speech possibly obscured the effects of lexical competition, or at least pupil dilation did not capture additive effects of lexical competition.

Even though results from fixations collected alongside the pupil dilation data are not the focus of this paper, these results correspond with the pupil dilation results: degradation also hampered the perception of mismatching cues, weakened and prolonged lexical competition. This suggests that pupil dilation might have captured lexical competition and sensitivity to mismatching cues if these two factors would have governed listeners' gazes. Taken together this implies that listening to degraded speech obscures the availability of acoustic cues, modifies the course of lexical competition, and increases the effort of listening.

The CI listeners showed a smaller response than the NS participants but their dilations increased due to lexical competition. This suggests that experience with degraded signals makes for a more natural course of lexical competition than short-term exposure to degradation, even though the sensitivity to mismatching cues is reduced.

The method of combining pupillometry with gaze fixations may provide a tool to objectively measure effort associated with difficulties in lexical processing or phonetic categorizations. It can compliment gaze fixations - an online measure of decision making - by adding a measure of the effort involved in the process of lexical or phonetic categorizations. This involuntary response of the perceptual system shows that ignoring the pan in panda does indeed require attentional resources, and is not as effortless as assumed even in natural speech.

7. REFERENCES

[1] Hoeks B., Levelt W. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods*, 25(1), 16–26.

[2] Iverson, P., Kuhl, P. K., AkahaneYamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for nonnative phonemes. *Cognition*, 87: B47-B57.

[3] Kahneman D., Beatty J. (1966). Pupil Diameter and Load on Memory. *Science*, 154, 1583-1585.

[4] Kuhl, P.K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, pp. 93-107.

[5] Salverda, AP, Dahan, D, & McQueen, JM (2004):The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension, *Cognition* 90 (1), 51-89.

[6] Shannon, R. Understanding hearing through deafness. (2007). Proc Natl Acad Sci,104(17):6883-4.

[7] Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1): 1-25.

[8] Zekveld, A.A., Heslenfeld, D.J., Johnsrude, I.S., Versfeld, N., & Kramer, S.E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage* ,101 (2014) 76–86.