



Benefit of Higher Maximum Force Output on Listening Effort in Bone-Anchored Hearing System Users

A Pupillometry Study

Bianchi, Federica; Wendt, Dorothea; Wassard, Christina; Maas, Patrick; Lunner, Thomas; Rosenbom, Tove; Holmberg, Marcus

Published in:
Ear and Hearing

Link to article, DOI:
[10.1097/AUD.0000000000000699](https://doi.org/10.1097/AUD.0000000000000699)

Publication date:
2019

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

[Link back to DTU Orbit](#)

Citation (APA):

Bianchi, F., Wendt, D., Wassard, C., Maas, P., Lunner, T., Rosenbom, T., & Holmberg, M. (2019). Benefit of Higher Maximum Force Output on Listening Effort in Bone-Anchored Hearing System Users: A Pupillometry Study. *Ear and Hearing*, 40(5), 1220-1232. <https://doi.org/10.1097/AUD.0000000000000699>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

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.

Benefit of Higher Maximum Force Output on Listening Effort in Bone-Anchored Hearing System Users: A Pupillometry Study

Federica Bianchi,¹ Dorothea Wendt,^{2,3} Christina Wassard,¹ Patrick Maas,^{1,4}
Thomas Lunner,^{2,3,5,6} Tove Rosenbom,¹ and Marcus Holmberg⁷

Objectives: The aim of this study was to compare listening effort, as estimated via pupillary response, during a speech-in-noise test in bone-anchored hearing system (BAHS) users wearing three different sound processors. The three processors, Ponto Pro (PP), Ponto 3 (P3), and Ponto 3 SuperPower (P3SP), differ in terms of maximum force output (MFO) and MFO algorithm. The hypothesis was that listeners would allocate lower listening effort with the P3SP than with the PP, as a consequence of a higher MFO and, hence, fewer saturation artifacts in the signal.

Design: Pupil dilations were recorded in 21 BAHS users with a conductive or mixed hearing loss, during a speech-in-noise test performed at positive signal-to-noise ratios (SNRs), where the speech and noise levels were individually adjusted to lead to 95% correct intelligibility with the PP. The listeners had to listen to a sentence in noise, retain it for 3 seconds and then repeat it, while an eye-tracking camera recorded their pupil dilation. The three sound processors were tested in random order with a single-blinded experimental design. Two conditions were performed at the same SNR: Condition 1, where the speech level was designed to saturate the PP but not the P3SP, and condition 2, where the overall sound level was decreased relative to condition 1 to reduce saturation artifacts.

Results: The P3SP led to higher speech intelligibility than the PP in both conditions, while the performance with the P3 did not differ from the performance with the PP and the P3SP. Pupil dilations were analyzed in terms of both peak pupil dilation (PPD) and overall pupil dilation via growth curve analysis (GCA). In condition 1, a significantly lower PPD, indicating a decrease in listening effort, was obtained with the P3SP relative to the PP. The PPD obtained with the P3 did not differ from the PPD obtained with the other two sound processors. In condition 2, no difference in PPD was observed across the three processors. The GCA revealed that the overall pupil dilation was significantly lower, in both conditions, with both the P3SP and the P3 relative to the PP, and, in condition 1, also with the P3SP relative to the P3.

Conclusions: The overall effort to process a moderate to loud speech signal was significantly reduced by using a sound processor with a higher MFO (P3SP and P3), as a consequence of fewer saturation artifacts. These findings suggest that sound processors with a higher MFO may help BAHS users in their everyday listening scenarios, in particular in noisy environments, by improving sound quality and, thus, decreasing the amount of cognitive resources utilized to process incoming speech sounds.

Key words: BAHS, Bone-anchored devices, Growth curve analysis, Listening effort, Maximum force output, Pupillometry, Pupil dilation, Speech-in-noise test.

(Ear & Hearing 2019;40:1220–1232)

INTRODUCTION

Speech intelligibility in complex listening scenarios can be challenging, especially in the presence of multiple talkers in a noisy environment. For some listeners, understanding speech in such a scenario is not only demanding, but it can also be exhausting. The concepts of fatigue, cognitive load, and listening effort have received increased attention in the past years (Pichora-Fuller & Kramer 2016; Pichora-Fuller et al. 2016), as it has become clear that the interplay between auditory and cognitive factors has a central role in everyday listening environments. According to the framework for understanding effortful listening (FUEL; Pichora-Fuller et al. 2016), listening effort is defined as the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a listening task. The need to consider listening effort when evaluating a listening situation has become of great importance, especially because physiological measures of effort, such as pupillometry, have been shown to provide additional information beyond behavioral performance (Zekveld et al. 2010; Wendt et al. 2017; Ohlenforst et al. 2018). Hence, considering both perceptual and cognitive factors during listening tasks can provide a powerful framework to observe complementary aspects of effortful listening.

It has been shown that people with hearing loss experience greater effort than normal-hearing (NH) listeners while performing a speech intelligibility task (Hick and Tharpe 2002; Hornsby 2013). Additionally, they seem to allocate cognitive resources differently than NH listeners, with a more prolonged and sustained effort as a function of signal-to-noise ratio (SNR) (Zekveld et al. 2011; Ohlenforst et al. 2017). Hearing-assistive devices can help people with hearing loss face every-day listening situations by means of amplification and advanced signal processing designed to reduce background noise and enhance the target speech signal. Indeed, a reduction in listening effort has been observed in listeners wearing hearing aids relative to the unaided condition (Downs

¹Oticon Medical AB, Kongebakken, Smørum, Denmark; ²Eriksholm Research Center, Oticon A/S, Rørtangvej, Snekkersten, Denmark; ³Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, Denmark; ⁴Oticon Medical, Karl-Wiechert-Allee, Hannover, Germany; ⁵Department of Behavioral Sciences and Learning, Linköping University, Sweden; ⁶Linnaeus Centre HEAD, The Swedish Institute for Disability Research, Linköping and Örebro Universities, Sweden; and ⁷Oticon Medical AB, Datavägen, Askim, Sweden.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

Copyright © 2019 The Authors. Ear & Hearing is published on behalf of the American Auditory Society, by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

1982; Hornsby 2013), as well as when a noise reduction scheme was applied (Sarampalis et al. 2009; Desjardins and Doherty 2014; Wendt et al. 2017; Ohlenforst et al. 2018).

While amplification and noise reduction algorithms in hearing aids were shown to reduce listening effort in people with hearing loss (Wendt et al. 2017; Ohlenforst et al. 2018), the effect of signal processing on listening effort with bone-anchored hearing systems (BAHS) has not yet been fully investigated. One previous study (Lunner et al. 2016) reported how the ability to recall words increased in patients with BAHS fitted on an abutment (percutaneous direct-drive device; Reinfeldt et al. (2015)) relative to the softband (skin-drive device). When the sound signal was transmitted directly to the bone via an abutment without skin dampening (i.e., direct drive), the listeners' recall ability increased. Because working memory has a limited capacity (Wingfield 2016), a trade-off occurs between resources allocated for processing the signal versus storing information. Thus, the findings of Lunner et al. (2016) suggest that the ability to remember words increased as a consequence of fewer cognitive resources being allocated to process speech with the direct-drive solution. It was hypothesized that the resources allocated to process speech decreased as a consequence of the higher signal quality and less distortions in the signal delivered via the direct-drive solution. This interpretation is in agreement with the evidence on hearing aids, where a relationship was also shown between working memory capacity and alterations of the speech signal due to hearing aid signal processing (Souza et al. 2015). However, to the knowledge of the authors, no study up to date has objectively investigated the link between signal fidelity and listening effort in BAHS users.

All BAHS have a low maximum force output (MFO), which is typically well below the listeners' loudness uncomfortable level (Zwartzkot et al. 2014). The maximum output of a modern bone anchored sound processor is controlled by signal processing and implemented as a limiting compressor (Lunner et al. 1997) that quickly attenuates the output signal in the frequency band where the MFO is reached, such that the physical saturation of the transducer never occurs. When the output signal reaches the MFO, the high compression ratio will introduce some artifacts, which may affect intelligibility and listening effort. These saturation artifacts in the signal are similar to those introduced by any general compression system (Stone and Moore 2007) and were shown to affect speech perception for hearing aid users (Dreschler 1988). However, the artifacts introduced by a limiting compressor are typically stronger due to the high compression ratio and fast attack times. Algorithms for managing the MFO can differ and can be implemented as a single-channel or multi-channel system, similarly as for other compression systems. The exact input level at which these artifacts appear not only depends on the MFO of the device, but also on the gain prescribed and, thus, on the patient's hearing loss. In BAHS, this may happen already at normal speech levels for broadband signals in patients with mixed hearing losses (Zwartzkot et al. 2014). Super power BAHS have a higher MFO to be suitable for listeners with a mixed conductive-sensorineural hearing loss up to 65 dB HL. The MFO of these devices will, thus, saturate at higher input levels and create fewer artifacts in the signal at medium speech levels. The aim of this study was to compare listening effort, as estimated via pupil dilation, during a speech intelligibility task performed with three BAHS sound

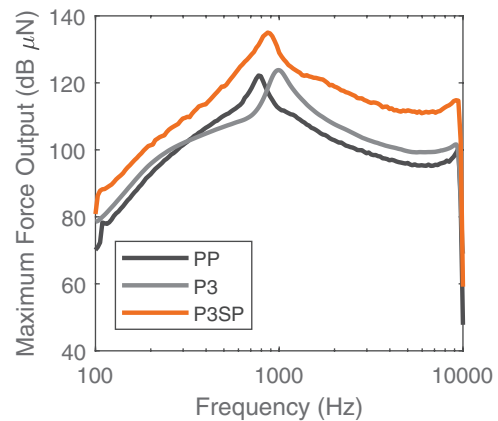


Fig. 1. Maximum force output (MFO) curves for a 90 dB SPL input at maximum gain for the three sound processors used in this study (Ponto Pro: PP; Ponto 3: P3; Ponto 3 SuperPower: P3SP).

processors (Oticon Medical AB, Askim, Sweden): Ponto Pro (PP), Ponto 3 (P3), and Ponto 3 SuperPower (P3SP). The three processors have a different MFO, as shown in Figure 1, with the PP delivering the lowest MFO, and the P3SP the highest. The P3SP and P3 only differ in the MFO they can deliver, but have the same MFO algorithm (multichannel MFO). The P3 and PP processors not only differ in the MFO they can deliver but also in the MFO algorithm, which is a single-channel system in PP and a multichannel system in the P3. The hypothesis was that the listener would allocate less cognitive resources to process the target speech when using the sound processor with the highest MFO (i.e., the P3SP), because fewer saturation artifacts would be present in the signal when compared with the PP. Importantly, the speech intelligibility task was performed at ecologically valid SNRs, which were individually adjusted to lead to 95% correct speech intelligibility with the PP. The speech levels were individually adjusted to saturate the PP (but not the P3SP) and corresponded to sound pressure levels (SPLs) that can be experienced by people with hearing loss during a conversation in noise (Wagener et al. 2008). Thus, listening effort was evaluated at SNRs and speech levels close to those that listeners with hearing loss are exposed to in noisy listening environments (Wagener et al. 2008; Smeds et al. 2015; Wu et al. 2018).

MATERIALS AND METHODS

Listeners

Twenty-one native Danish BAHS users (8 males and 13 females) with a conductive or mixed-conductive sensorineural hearing loss were recruited from the Oticon Medical test person database. The listeners were between 20 and 80 years old (mean 58.8 ± 17.0 years; Table 1). The listeners were all experienced BAHS users (at least 6 months of experience with Ponto) except for one listener who had only 4 months of experience with Ponto. Prior to the commencement of the study, 18 listeners were already unilaterally fitted with one Ponto device on abutment and three listeners were bilaterally fitted with two Ponto devices on abutment. During the study, all participants were unilaterally fitted with specific nominal devices on one selected ear (the tested ear is listed in Table 1), and the listeners' own device was not used. For the 18 unilaterally fitted patients, the selected ear was the one with the abutment, while for the three bilaterally

TABLE 1. Characteristics of the 21 BAHS users and individual presentation levels

Listener	Age (years)	Ear tested	PTA (dB HL)	Speech Level (dB SPL)	Noise Level (dB SPL)	SNR (dB)
1	58	Right	15.0	80	71.9	8.1
2	66	Right	30.0	77	64.3	12.7
3	68	Right	12.5	78	71.8	6.2
4	70	Right	37.5	70	55.9	14.1
5	70	Left	42.5	71	52.2	18.8
6	29	Right	20.0	79	73.8	5.2
7	59	Left	17.5	78	70.5	7.5
8	80	Right	40.0	71	58.0	13.0
9	62	Right	13.6	79	72.6	6.4
10	62	Right	31.3	77	68.8	8.2
11	20	Right	32.5	75	70.7	4.3
12	29	Right	43.8	68	52.8	15.2
13	49	Left	18.8	79	74.5	4.5
14	80	Left	26.3	80	70.6	9.4
15	76	Left	38.8	73	66.2	6.8
16	64	Left	28.8	79	73.0	6.0
17	58	Right	31.3	76	72.3	3.7
18	74	Right	28.8	76	71.2	4.8
19	40	Left	15.0	79	75.0	4.0
20	45	Left	22.5	78	70.4	7.6
21	75	Left	36.3	75	60.6	14.4

The PTA refers to the mean BC in situ thresholds at 0.5, 1, 2, and 3 kHz. The speech and noise levels are the individual levels used in condition 1. The speech and noise levels used in condition 2 were decreased by 5 dB relative to the levels of condition 1; the SNR was, thus, the same in both conditions.

BAHS, bone-anchored hearing system; BC, bone conduction; PTA, pure-tone average; SNR, signal-to-noise ratio.

fitted patients, only the ear with the best (lowest) pure-tone average (PTA), as measured in situ with the sound processor (0.5, 1, 2, and 3 kHz), was fitted. For all listeners, the opposite ear was occluded by means of an earplug to ensure that perception primarily happened through the test ear and the corresponding device. The individual bone conduction (BC) hearing thresholds, as measured in situ via the sound processor (before applying the gain), are depicted in Figure 2 (individual thresholds in gray; mean threshold in black). For all listeners, the PTA (Table 1), calculated for the BC in situ thresholds at 0.5, 1, 2, and 3 kHz, was within the fitting range of all three sound processors (≤ 45 dB HL). This study was reviewed by the Ethics Committee of the Capital Region of Denmark (Protocol number: H-17021915).

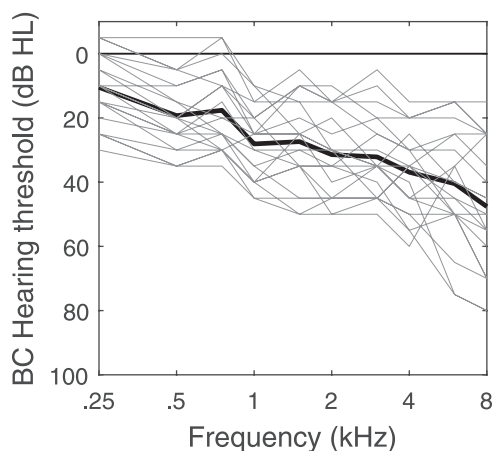


Fig. 2. Bone conduction (BC) hearing thresholds of the 21 listeners included in the study (in gray), as measured in situ via the Ponto 3 SuperPower (P3SP). The mean BC threshold is depicted by the black curve.

Devices and Specifications

Three BAHS sound processors were used: PP, P3, and P3SP (Oticon Medical AB, Askim, Sweden). The PP was released in 2009, while both the P3 and the P3SP have been commercially available since 2016. The three devices differ in terms of MFO, with the P3SP delivering the highest MFO (Fig. 1), and in the number of MFO frequency bands. The PP has a low MFO, leading to a low dynamic range especially at low frequencies, and a one-channel MFO. Hence, when the output signal reaches the MFO of the PP (already for relatively low input levels at low frequencies), the output of the device is attenuated in the entire frequency range, leading to saturation artifacts in the signal. The P3 and the P3SP have, instead, a 4-channel MFO with the following cutoff frequencies: 200–625 Hz, 625–1562.5 Hz, 1562.5–3125 Hz, and 3125–9500 Hz. Hence, if the lowest frequency band is attenuated due to MFO saturation, the other three bands are still performing with full dynamics leading to fewer saturation artifacts in the output signal. Hence, the combination of the MFO limit (Fig. 1), together with the number of MFO frequency bands, will determine the amount of saturation artifacts for a given input signal and a given gain setting. A comparison between the P3SP and P3 will reflect the pure effect of the MFO limit, while a comparison between P3SP and PP will reflect the effect of both the MFO limit and MFO frequency bands. A comparison between P3 and PP will also reflect both factors, although it will be mostly dominated by the number of MFO frequency bands.

During fitting, noise reduction was turned off, directionality settings were in omnidirectional mode, and the volume and mute settings were off. No fine tuning was performed to ensure that the difference between the sound processors was as much as possible confined to the difference in MFO level and MFO frequency bands.

Experimental Procedure

Each listener carried out two sessions, each of 2 hours in duration. At the first visit, feedback measurement and BC in-situ threshold testing were performed with the P3SP sound processor. The gain settings were prescribed according to the modified NAL-NL1 prescription used in Ponto. The gain settings from the P3SP were then transferred to the other two sound processors, the PP and the P3, to ensure that the gain settings were similar across devices. However, because the maximum gain that can be applied is lower for the PP, four subjects had a lower gain with the PP (relative to the P3SP) at 4 kHz (2 to 4 dB difference for listeners 5, 12, 15, and 21) at the input sound levels used in this study. The gain at the other frequencies was not affected, except for listener 12 (1 dB lower gain with the PP at 1 kHz). The P3 has also a lower maximum gain than the P3SP, but this did not affect the gain settings, which were thus the same for the P3SP and P3 at the presentation levels used in this study.

After the fitting, the listeners performed a speech intelligibility test with Danish HINT sentences (Nielsen and Dau 2011) to measure the individual SNR corresponding to 50% (SRT50) and 80% (SRT80) correct performance with the PP. Three lists (20 sentences per list) were carried out: the first list was a training list to measure the SRT50, the second list was the test list to measure the SRT50, and the third list was the test list to measure the SRT80. During these measurements, the speech signal was fixed at an individual level for each listener to ensure that the P3SP would not reach saturation in its second MFO band (see “Speech Level” section for further details; the individual levels are listed in Table 1). The noise level was adjusted in 0.8 dB steps, depending on how many words were correctly repeated: from a decrease of 3.2 dB for 0 words correct to an increase of 0.8 dB for 5 words correct in the SRT80 measurement; from a decrease of 2 dB for 0 words correct to an increase of 2 dB for 5 words correct in the SRT50 measurement. The step size was doubled for the first four sentences. A psychometric function was fit to the SRT50 and SRT80 to estimate the SRT95, that is, the SNR that would lead to 95% correct performance with the PP. The PP was used as a baseline, to ensure that 95% correct was reached even when the processor was in saturation. The noise levels estimated for each listener for SRT95 are listed in Table 1.

At the second visit, pupil dilations were recorded during a speech intelligibility test with HINT sentences presented at the fixed SRT95 (the individual speech and noise levels are listed in Table 1). The listeners were instructed to look at the camera positioned in front of them, listen to each sentence, and repeat it back once the background noise stopped. After a training list with PP, all three sound processors were tested at the same SRT95 with a single-blinded randomized procedure (1 list of 25 sentences per processor). This condition is referred to as condition 1.

After a break, the test was carried out at a lower overall level (condition 2), where both the speech and noise levels were decreased by 5 dB relative to condition 1. Condition 2 was, hence, presented at the same SNR as in condition 1 but was assumed to generate fewer artifacts in the signal relative to condition 1.

Speech Level

Because the output of the sound processor depends on the individual gain settings, the target speech level needed to be individually adjusted to compare listening effort with a device

that was in saturation (PP) relative to a device that was not in saturation (P3SP). In the fitting software (Genie Medical 2016.1; Oticon Medical AB, Askim, Sweden), the stationary response of the sound processor can be accurately simulated for different types of input signals (e.g., pure tones, warble tones, white noise, ICRA stationary signal) at different SPLs ranging from 45 to 90 dB. To simulate the output curves for the HINT speech material, the long-term average speech spectrum (LTASS) of the ICRA signal was chosen (ANSI S3.5), because this was the closest simulation to the LTASS of the HINT material. For each listener, the output LTASS was, first, simulated for the individual gain settings in the fitting software. On the basis of these simulated curves, the input speech level was then chosen to generate an output that would not saturate the MFO of the P3SP. Because the speech signal has a dynamic range of 30 dB, +12 dB, and –18 dB, around the average presentation level (Byrne et al. 1994; Seewald et al. 2005), the individual input speech level was adjusted to lead to an output level of 12 dB below the MFO of the P3SP at 750 Hz*. Hence, for each listener, the input speech level was adjusted to avoid saturation artifacts in the second MFO band of the P3SP. For each listener, this speech level (listed in Table 1) was kept fixed throughout all measurements of condition 1, and lowered by 5 dB in condition 2. In contrast, the whole MFO frequency band of the PP was modeled to be in saturation at this input level, leading to attenuation and artifacts in the output signal across the whole frequency range. It should be noted that the mean speech levels, averaged across all listeners, were 76 dB SPL in condition 1 and 71 dB SPL in condition 2, corresponding to a loud speech signal but not shouted speech (Pearsons et al. 1977). These values are within the 75th percentile of the distribution of SPLs encountered by people with hearing loss during conversations in noise (Wagener et al. 2008).

The sensation level (SL) of the output speech signal was calculated, for each listener and processor based on the FLogram representation (Hodgetts and Scollie 2017; Scollie et al. 2018) in the fitting software. The SL was calculated as the difference between the simulated output force level (dB μ N) for the input speech levels in conditions 1 and 2 (Table 1) and the BC in situ threshold converted to force level, averaged between 0.5, 1, 2, and 3 kHz.

Experimental Set Up

The experiment performed in this study was carried out in a soundproof booth, with a similar set up as in Wendt et al. (2017). Five loudspeakers (Genelec 8040A; Finland) were placed in a circumference with a 120-cm radius at 0°, ±90°, and ±150° azimuth. The listener was sitting in the center. An eye-tracking camera was placed in front of the listener (0° azimuth) at a distance of approximately 60 cm from the eyes. The eye tracker system was the iView X RED System (SensoMotoric Instruments from Teltow, Germany). Pupil dilation was recorded at a sampling frequency of 60 Hz.

* Due to the low dynamic range of all three devices under investigation, the first MFO band was always expected to be in saturation on both the P3 and the P3SP, for the gain settings and presentation levels utilized in the study. Therefore, the second MFO band was chosen for the speech-level calculation, more specifically the 750-Hz trimmer band, because this was the lowest available signal processing band within the second MFO band.

For each processor and condition, a list of 25 Danish HINT sentences (Nielsen and Dau 2011) was presented in a four-talker babble background noise (two male and two female speakers reading text from a newspaper; Wendt et al. 2017). The audio files of the four-talker babble had the same long-term average frequency spectrum as the Danish HINT sentences. The target speech signal was presented from the loudspeaker at the 0° azimuth position. Each of the four babble talkers was presented from one of the four loudspeakers at ±90° and ±150° azimuth. One male and one female talkers were always presented at ±90°, but the female versus male position switched from training to testing, and from condition 1 to condition 2, to have pitch-balanced conditions.

Each trial started with 3 seconds of noise (four-talker babble), followed by the HINT sentence with a mean duration of 1.5 seconds. Thus, the sentence ended on average 4.5 seconds after trial onset. The noise was presented throughout the playback of the sentence and continued for 3 seconds after the presentation of the sentence, that is, until, on average, 7.5 seconds after trial onset. After retaining the sentence for 3 seconds (“response preparation window”), the listeners were instructed to repeat the sentence after the offset of the noise (“response window”). The number of words correctly repeated for each sentence was used as scoring method. The presentation of the stimuli was controlled by a PC using MATLAB (MathWorks, Natick, MA) based programming.

Pupil Data Processing

The individual pupil data were processed similarly as in (Wendt et al. 2017). For each participant and condition, the mean pupil dilation was calculated for each sentence. Dilations exceeding twice the standard deviation relative to the mean were considered as eye blinks. Trials with more than 15% of eye blinks were disregarded from further analysis. For the remaining sentences, the eye blinks were removed using linear interpolation from 35 to 75 ms preceding and following the blink, respectively. The data were then filtered through a 35-point moving average smoothing filter to remove high-frequency artifacts. For each sentence, a baseline value was computed as the mean pupil diameter recorded during the 1-second time range preceding the sentence onset (i.e., during the noise presentation). This baseline value was subtracted from each pupil curve to obtain a baseline-corrected pupil dilation. All baseline-corrected pupil curves obtained for sentences between the 6th and 25th were averaged to obtain the mean dilation for each participant and condition. The first five sentences were disregarded to allow for the participant to adjust to the task.

An additional analysis comparing pupil dilation for correct versus incorrect sentences was performed including all 25 sentences. The pupil responses recorded for all sentences that were correctly repeated (100% words correct) were averaged and compared with the pupil responses for sentences with at least one word that was not correctly repeated (<100% correct).

The peak pupil dilation (PPD) was obtained for each subject as the maximum dilation between 3 and 6 seconds after the trial onset (i.e., for the whole sentence presentation until about 1.5 seconds after the sentence offset). After manual inspection of the individual peaks, the time range to calculate the PPD was restricted to start at 4 seconds after trial onset for four subjects, who had an initial decay instead of a dilation at stimulus onset resulting in a delayed peak. Two of these

four subjects were excluded from the analysis of condition 2, due to a pupillary decrease in size after stimulus onset, instead of dilation (relative to baseline). This is the case when a stimulus-unrelated dilation occurs during the baseline window and, hence, no stimulus-related dilation can be recorded after baseline (and no PPD can be individuated).

Statistical Analysis

The analysis on the PPD was performed using planned comparisons (paired one-tailed *t*-tests), based on the a priori hypothesis that, in condition 1, the PPD obtained with the P3SP would be lower than the one obtained with both the PP and P3, as well as the PPD obtained with the P3 would be lower than the one for PP. No a priori hypothesis was formulated for condition 2, hence paired two-tailed *t*-tests were performed.

Mixed-linear models were implemented in R-studio using the statistical package lmerTest (Kuznetsova et al. 2017) to further analyze both the behavioral and pupillometry data.

A mixed-linear model analysis of variance was fit to the behavioral data to analyze the performance in the HINT test. Processor (three levels) and condition (two levels) were used as fixed factors, while listener was considered as a random effect. Post hoc analysis was carried out via contrasts of least-square means, and the *p*-values were corrected for multiple comparisons (by *n* = 3 processors) using the Tukey method.

Growth Curve Analysis (0–5 seconds After Sentence Onset) • Growth curve analysis (GCA; Mirman 2014) was used to model the changes in pupil dilation over time (Kuchinsky et al. 2013; Winn et al. 2015; Winn 2016; Wendt et al. 2018). A third-order polynomial function was fit to the pupil dilation data in the time range from sentence onset (3 seconds after trial onset) until about 3.5 seconds after sentence offset (8 seconds after trial onset). Orthogonal polynomial time terms were used to make the time vectors independent, such that the parameter estimates could be interpreted independently. The polynomial function was in the form of a mixed model, with processor as a fixed factor, as well as the interaction of processor with each of the polynomial time terms. The formula used was the following:

$$\text{Pupil Dilation} = (\text{Linear} + \text{Quadratic} + \text{Cubic}) \\ \times \text{Processor} + (1 + \text{Linear} + \text{Quadratic} + \text{Cubic} | \text{Subject}),$$

where the Linear, Quadratic, and Cubic are the orthogonal terms; × indicates the interaction with Processor, and the terms reported in the second parenthesis are the random factors. The random factors included the effects of listeners, and of each of the time terms, to account for the variability in the time course of dilation across listeners. A stepwise backwards elimination procedure did not suggest the elimination of any term in the model (the *p*-values for the fixed effects were calculated based on Satterthwaite’s method; *p*-values for the random effects were based on the likelihood ratio test). This model allowed for the statistical comparison of overall level of effort (the intercept term, similar to the concept of “area under the curve”; Mirman et al. 2008; Winn et al. 2015), rate of growth/decay (the linear term), changes in the rate of growth/decay (quadratic term), and the steepness of the curve around inflection points (cubic term; Mirman et al. 2008), with the three different sound processors. Because effects on terms higher than the quadratic can be difficult to interpret (Mirman et al. 2008), this study focused on the effect of processor on the

intercept, linear, and quadratic terms. The hypothesis was that the overall effort for listening and preparing the response (intercept term) with the P3SP would be lower than the overall effort with both the P3 (pure effect of MFO level) and PP (effect of MFO level and number of MFO frequency bands).

Growth Curve Analysis (2–5seconds After Sentence Onset) • A GCA was additionally performed during the response preparation period, that is, the time interval where the decay of the pupillary curve can be observed. This time window ranged from 0.5 seconds after sentence offset (i.e., 5 seconds after trial onset) until about 3.5 seconds after sentence offset (i.e., 8 seconds after trial onset; 0.5 seconds after the response window onset). It is argued that this time interval reflects the cumulative memory load (Piquado et al. 2010) and more cognitive aspect of the task during the retention interval (involving rehearsing or reconstructing the sentence), as opposed to the listening effort during sentence presentation, which mostly reflects differences in the perception of the input signal (Winn et al. 2015). Hence, this second GCA allowed for analyzing (1) differences across processors in pupil dilation during sentence rehearsal and reconstruction (intercept term), (2) effort release during response preparation (linear term). It was hypothesized that a higher effort release, as indicated by a lower overall effort in the response preparation window (intercept term) and/or a steeper negative slope (linear term), would be obtained with the P3SP relative to the other two processors. The same third-order polynomial function was used for this GCA, because it showed an improved fit relative to a simpler model (second-order polynomial function). The fit was improved based on a χ^2 test ($p < 0.0001$) and a reduction of the Akaike Information Criterion (AIC), which is an indicator of the goodness-of-fit of a model.

RESULTS

Behavioral Performance During HINT

The mean performance in the speech intelligibility task (i.e., percentage of words correctly repeated) is depicted in Figure 3 for condition 1 (Fig. 3A) and condition 2 (Fig. 3B). The mean performance with the three processors was 92.3%, 94.3%, and 95.6% in condition 1; 89.6%, 91.0%, and 93.4% in condition 2, with the PP, P3, and P3SP, respectively. The analysis of variance revealed a significant effect of both fixed factors: processor [$F(2, 100) = 5.33; p = 0.006$] and condition

[$F(1, 100) = 9.29; p = 0.003$]. The interaction of processor and condition was not significant, indicating a similar effect of processor in the two conditions. The post hoc analysis showed that the effect of processor was significant only between the PP and the P3SP ($p = 0.004$, after Tukey correction for multiple comparisons). No significant differences in performance were observed between the PP and the P3 ($p = 0.281$) and between the P3 and the P3SP ($p = 0.199$). The improvement in performance obtained with the P3SP relative to the PP was not correlated with the listeners' PTA (Spearman correlation; condition 1: $\rho = 0.09, p = 0.708$; condition 2: $\rho = 0.39, p = 0.082$). This finding confirms that the improvement with the P3SP (re PP) was not related to gain differences and suggests that a higher MFO and, thus, fewer saturation artifacts in the speech signal could improve intelligibility in most listeners, independent of their hearing loss.

Figure 4 shows the correlation between the individual behavioral performance with the three sound processors (% words correct) and the PTA, in condition 1 (left panel) and condition 2 (right panel). A significant correlation was obtained between speech intelligibility performance and the PTA only in condition 2, with all three processors (Spearman correlation; PP: $\rho = -0.71, p < 0.001$; P3: $\rho = -0.60, p = 0.004$; P3SP: $\rho = -0.72, p < 0.001$). This finding suggests that performance in condition 2 was limited by audibility with all three processors (as further addressed in the Discussion). No significant correlation was found in condition 1 (Spearman correlation; PP: $\rho = -0.25, p = 0.265$; P3: $\rho = -0.43, p = 0.052$; P3SP: $\rho = -0.34, p = 0.137$). A correlation was also carried out between behavioral performance and speech sensation level (SL; Fig. 1, Supplemental Digital Content 1, <http://links.lww.com/EANDH/A506>). Similarly as with PTA, a significant correlation was obtained between speech intelligibility performance and SL only in condition 2 (Spearman correlation; PP: $\rho = 0.70, p < 0.001$; P3: $\rho = 0.59, p = 0.005$; P3SP: $\rho = 0.69, p = 0.001$), indicating a worsening in performance with decreasing SL.

Pupillometry

The left panels in Figure 5 depict the mean pupil dilation during the speech intelligibility task (Fig. 5A: condition 1; Fig. 5B: condition 2), normalized relative to the 1 second of

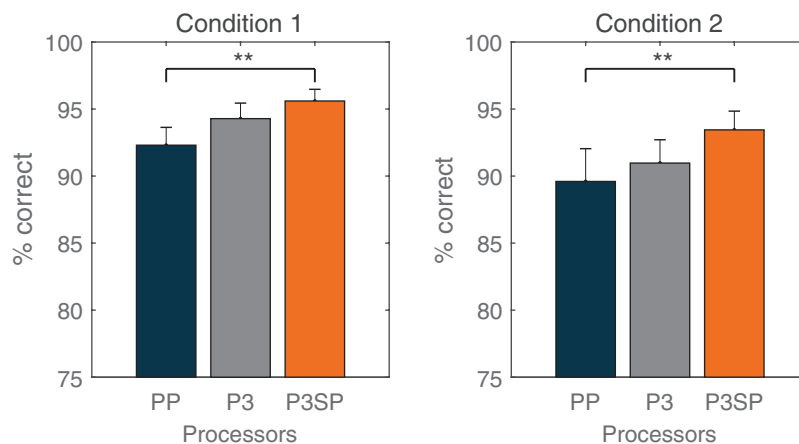


Fig. 3. Mean performance (% words correctly repeated) in the speech intelligibility test (N = 21 listeners) with the three sound processors (Ponto Pro: PP; Ponto 3: P3; Ponto 3 SuperPower: P3SP), in condition 1 (left panel) and condition 2 (right panel). Error bars depict the standard error of the mean.

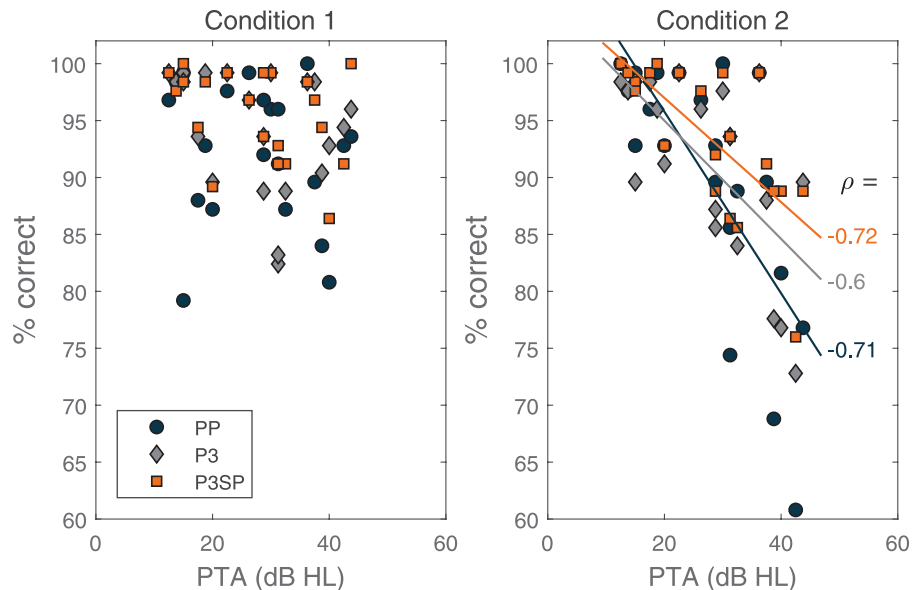


Fig. 4. Speech intelligibility (% words correct) with Ponto Pro (PP), Ponto 3 (P3), and Ponto 3 SuperPower (P3SP) for the individual listeners ($N = 21$), as a function of their pure-tone average (mean bone conduction hearing thresholds at 0.5, 1, 2, and 3 kHz), for condition 1 (left panel) and condition 2 (right panel). Linear regression lines are depicted for visualization purposes when the correlation was significant ($p < 0.05$; Spearman correlation coefficient, ρ , is also reported next to the regression line).

noise preceding sentence onset (baseline from -1 to 0 seconds). Relative to the baseline, the pupil dilated during sentence presentation until reaching a maximum value of dilation at about 2 seconds after sentence onset (i.e., at about 0.5 seconds after sentence offset). After the peak dilation, the pupil decreased in size at different decay rates depending on the sound processor used. During the response window, the pupil dilated again.

The right panels in Figure 5 depict the mean PPD of the first peak, for the three sound processors. In condition 1, which was tested at a speech level individually adjusted to saturate the PP but not the P3SP, the PPD obtained with the P3SP was significantly lower than the PPD of the PP (paired-sample one-tailed t -test; $p = 0.023$). No difference in PPD was obtained between the P3 and the P3SP (paired-sample one-tailed t -test; $p = 0.455$), or between the P3 and PP (paired-sample one-tailed t -test; $p = 0.095$). In condition 2, which was tested at a lower overall level to generate fewer saturation artifacts, no difference in PPD was observed across sound processors (paired-sample two-tailed t -tests; $p > 0.05$). Further analysis of the pupil dilation was performed via the GCA (see next sections).

GCA (0–5 seconds After Sentence Onset)

A visual comparison of the GCA and the modeled pupil responses is presented in Figure 6A, where the polynomial growth curve model is depicted in the left panel for condition 1 and in the right panel for condition 2, together with the mean pupil responses, sampled every 200 ms between 0 and 5 seconds after sentence onset. Table 2 presents a summary of the intercept (overall dilation in the 0–5 seconds window comprising listening and response preparation), linear (overall slope), and quadratic (changes in the slope) terms for each of the three processors in conditions 1 and 2. The GCA model fit and full output summary are presented in the Supplemental Digital Content 2, <http://links.lww.com/EANDH/A507> (Tables 1 and 2, for condition 1 and condition 2, respectively). In condition 1, the overall pupil

dilation, indicated by the intercept term (“area under the curve”), was of 0.055, 0.034, and 0.028 mm, for the PP, P3, and P3SP, respectively (Table 2). The overall dilation obtained with the P3 and P3SP was significantly lower than the one with the PP ($p < 0.0001$; Table 1 in Supplemental Digital Content 2, <http://links.lww.com/EANDH/A507>). Comparisons between the P3SP and the P3 were obtained by changing the reference in the model to the P3SP. Significant differences in overall dilation were also observed between these two devices, with the P3SP showing a significantly lower overall dilation than the P3 ($p < 0.0001$). The overall slope (linear term) for the P3 and P3SP was significantly shallower relative to the one of the PP ($p < 0.0001$). The effect of processor was also significant on the quadratic term, indicating slower changes in the rate of growth/decay for the P3 and P3SP (i.e., a less pronounced downwards/negative curvature) relative to the PP, and on the cubic term, indicating a shallower slope around the peak of the response for the P3 and P3SP relative to the PP. No significant differences in the linear, quadratic, and cubic terms were observed between the P3 and the P3SP.

In condition 2 (right panel in Fig. 6A, Table 2, and Table 2 in Supplemental Digital Content 2, <http://links.lww.com/EANDH/A507>), the effect of processor was significant on the intercept term (“area under the curve”), with the P3 and P3SP showing lower overall pupil dilations than the PP ($p < 0.0001$). The P3 and P3SP also showed a significant decrease in overall slope (linear term) relative to the PP ($p < 0.0001$). When changing the reference in the model to the P3SP, significant differences were observed between P3 and P3SP, with the P3 showing lower overall dilation than the P3SP ($p < 0.0001$) and less negative overall slope ($p < 0.001$). The effect of processor was significant also on the quadratic term, indicating faster changes in the overall slope for the P3 and P3SP (i.e., a more pronounced downwards/negative curvature) relative to the PP, and on the cubic term (only for P3SP relative to PP).

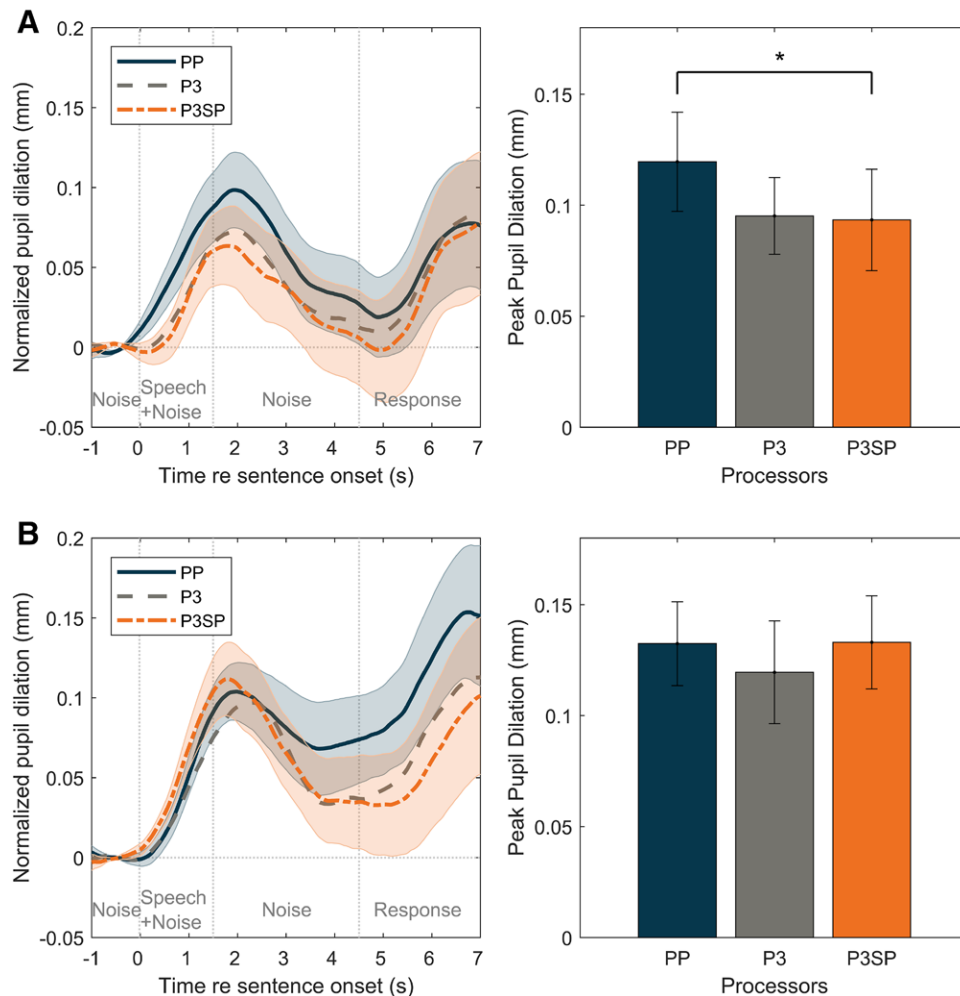


Fig. 5. Mean pupil responses averaged across participants, for condition 1 (panel A; $N = 21$ listeners) and condition 2 (panel B; $N = 19$ listeners). Left panels: Time course of the mean pupil dilation during the speech intelligibility test with the three sound processors [solid blue line: Ponto Pro (PP), dashed gray line: Ponto 3 (P3), dash-dotted orange line: Ponto 3 SuperPower (P3SP)]. The shaded area depicts the standard error of the mean. For readability, the shaded area of the P3 is not shown. Right panels: Mean peak pupil dilation (PPD), calculated as the average PPD across listeners. Error bars depict the standard error of the mean.

GCA (2–5 seconds After Sentence Onset)

A separate GCA was carried out on the time window between the sentence offset and the response window to specifically compare the mean value of dilation during decay (intercept term), as well as the decay rates (linear term) between the three processors during response preparation (i.e., sentence rehearsal/reconstruction). Figure 6B depicts a visual comparison of the GCA and the modeled pupil responses for condition 1 (left panel) and condition 2 (right panel). Table 3 presents a summary of the intercept (overall dilation) and linear terms (slope of decay) for each of the three processors in conditions 1 and 2. In condition 1, all three sound processors had a negative slope (linear term) significantly different from zero, indicating a decay toward baseline. However, only the P3 and P3SP showed a significant negative slope of decay in condition 2, while the slope of the PP did not significantly differ from zero indicating no release of effort during response preparation.

The full GCA output summary is presented in the Supplemental Digital Content 2, <http://links.lww.com/EANDH/A507>

(Tables 3 and 4 for condition 1 and condition 2, respectively). In condition 1, the effect of processor was significant on the intercept term, as well as on the linear term. Specifically, the P3 and P3SP showed a significantly smaller dilation, as indicated by the intercept term, during response preparation than the PP ($p < 0.0001$) and a slower rate of decay (linear term; $p < 0.0001$). When changing the reference in the model to the P3SP, a significant lower overall dilation was found with the P3SP relative to the P3 (intercept; $p < 0.001$), while no difference was present between P3SP and P3 in the rate of decay (linear term; $p = 0.573$).

In condition 2, the P3 and P3SP also showed a significantly smaller dilation in the decay window than the PP (intercept; $p < 0.0001$) and a faster rate of decay (linear term; $p < 0.0001$). Changing the reference in the model to the P3SP revealed no significant difference in overall dilation between P3SP and P3 (intercept; $p = 0.547$), and no difference in the rate of decay (linear term; $p = 0.232$).

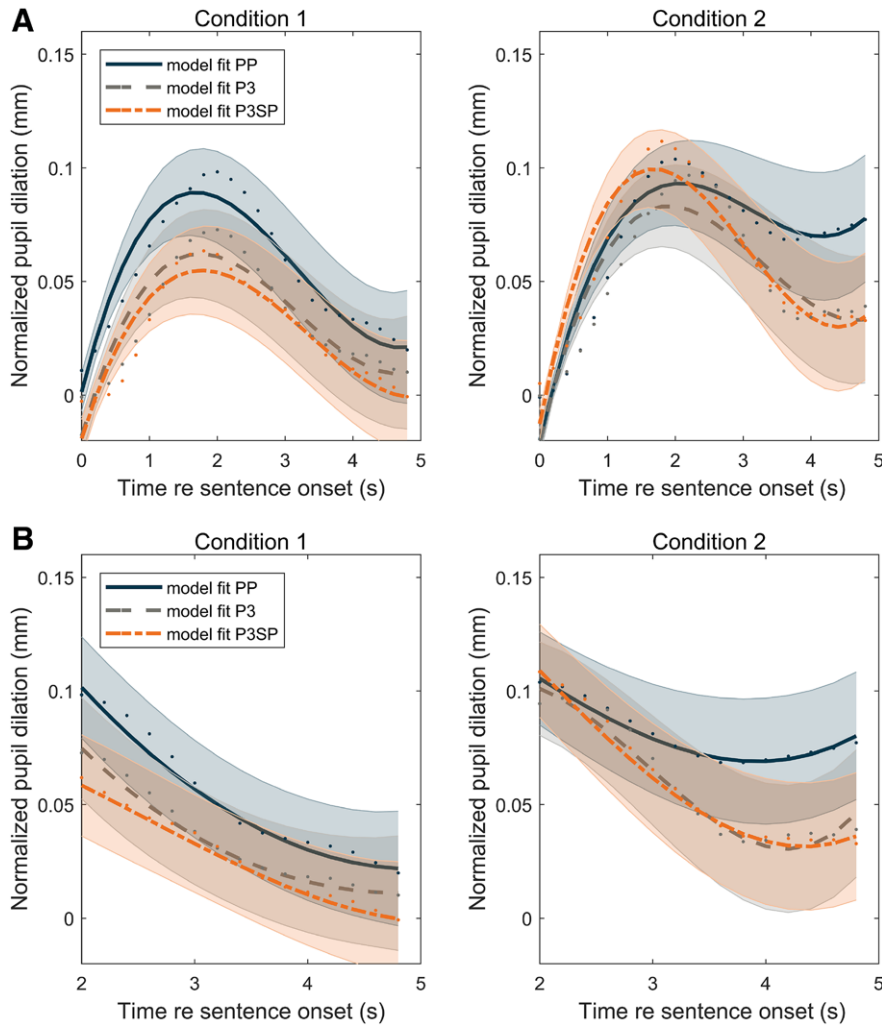


Fig. 6. Growth curve analysis (GCA) model fit [solid blue line: Ponto Pro (PP), dashed gray line: Ponto 3 (P3), dash-dotted orange line: Ponto 3 SuperPower (P3SP)] overlaid to the normalized pupil dilation (depicted by the dotted symbols every 200-ms time bin). A, GCA from sentence onset until 3.5 seconds after sentence offset, for condition 1 (left panel) and condition 2 (right panel). B, GCA during response preparation (from 0.5 seconds until 3.5 seconds after sentence offset) for condition 1 (left panel) and condition 2 (right panel). The shaded area depicts the standard error of the model. The output summary of the GCA can be found in Tables 2 and 3.

Effort for Processing Correct Versus Incorrect Sentences

The pupil responses were analyzed separately for all sentences that were correctly repeated (100% words correct) and sentences with at least one word that was not correctly repeated

(<100% correct). Figure 7 shows the pupil dilations averaged across all correct sentences on the left panels and the incorrect sentences on the right panels (Fig. 7A: condition 1; Fig. 7B: condition 2). In condition 1, there were, on average, 5, 4, and 3

TABLE 2. Growth curve analysis output summary of the intercept, linear, and quadratic terms for each sound processor in the listening and response preparation windows (0–5 seconds relative to sentence onset), for conditions 1 and 2

	PP		P3		P3SP	
	Estimate	<i>p</i>	Estimate	<i>p</i>	Estimate	<i>p</i>
Condition 1						
Intercept	0.055	0.004**	0.034	0.056	0.028	0.114
Linear	−0.043	0.199	−0.021	0.528	−0.026	0.439
Quadratic	−0.110	<0.0001***	−0.096	0.001**	−0.095	0.001**
Condition 2						
Intercept	0.069	<0.0001***	0.053	0.005**	0.060	0.002**
Linear	0.073	0.163	0.012	0.814	−0.029	0.570
Quadratic	−0.102	0.001**	−0.117	<0.0001***	−0.124	<0.0001***

p < 0.05 indicates an estimate significantly different from 0 (baseline); (**p* < 0.05; ***p* < 0.01; ****p* < 0.001). PP, Ponto Pro; P3, Ponto 3; P3SP, Ponto 3 SuperPower.

TABLE 3. Growth curve analysis output summary of the intercept and linear terms for each sound processor in the response preparation window (2–5 seconds relative to sentence onset), for conditions 1 and 2

	PP		P3		P3SP	
	Estimate	<i>p</i>	Estimate	<i>p</i>	Estimate	<i>p</i>
Condition 1						
Intercept	0.051	0.024*	0.032	0.133	0.025	0.237
Linear	-0.095	<0.0001***	-0.075	0.003**	-0.073	0.004**
Condition 2						
Intercept	0.080	0.002*	0.058	0.021*	0.057	0.023*
Linear	-0.034	0.286	-0.086	0.012*	-0.093	0.007**

p < 0.05 indicate an estimate significantly different from 0 (baseline). (**p* < 0.05; ***p* < 0.01; ****p* < 0.001). PP, Ponto Pro; P3, Ponto 3; P3SP, Ponto 3 SuperPower.

sentences per listener that were not correctly repeated out of a list of 25 sentences with the PP, P3, and P3SP, respectively. In condition 2, there were 6, 5, and 4 incorrect sentences per list with the PP, P3, and P3SP, respectively. Although the standard error of the mean was quite large with such few incorrect sentences, the pattern of pupil dilation for incorrect sentences was

consistent between conditions 1 and 2, showing a sustained effort that prolonged until the response window.

DISCUSSION

The behavioral outcomes of the current study showed how speech intelligibility in background babble noise significantly

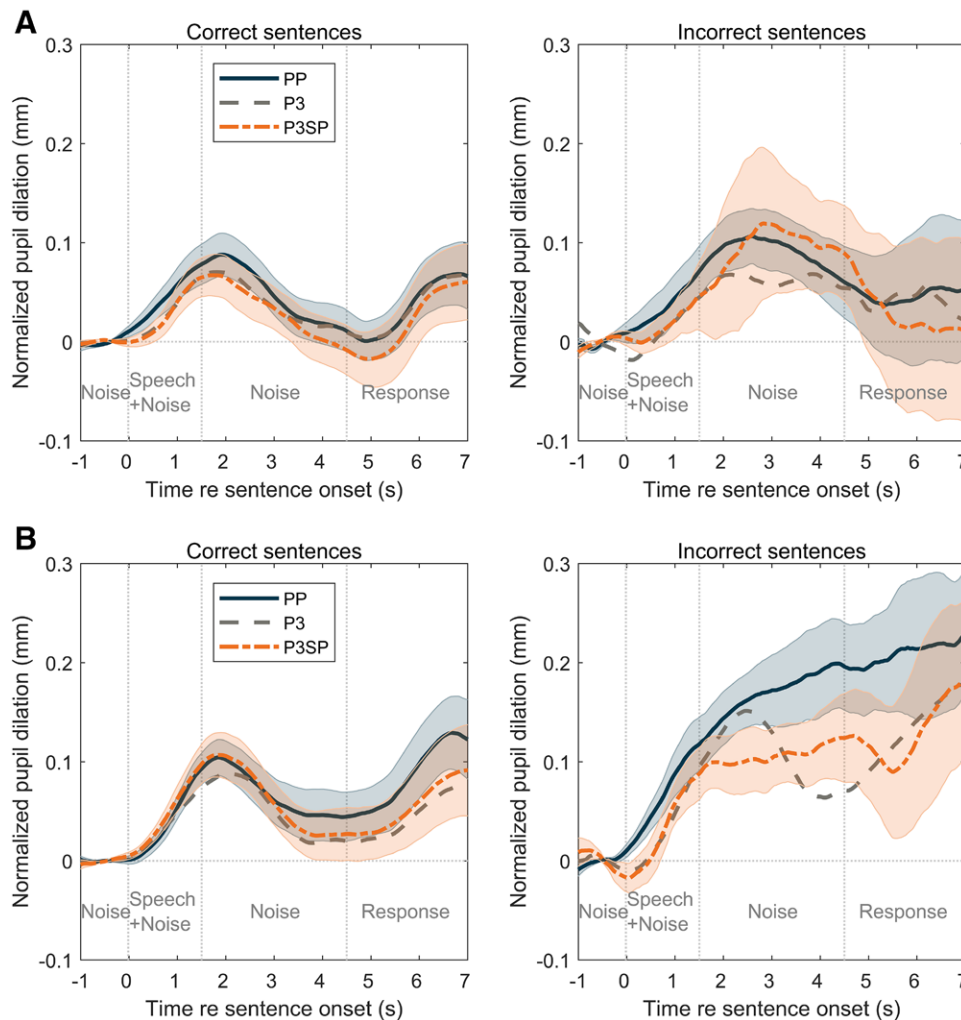


Fig. 7. Time course of the mean pupil dilation, for sentences with 100% of words correctly repeated (left panels) and sentences with <100% words correctly repeated (right panels; at least one incorrect word). The shaded area depicts the standard error of the mean. For readability, the shaded area of the Ponto 3 (P3) is not shown. A, Condition 1 (N = 21 listeners). B, Condition 2 (N = 19 listeners).

improved when the listeners were wearing the P3SP as compared with the PP, despite the fact that the gain settings were, for each listener, similar across devices. The performance with the P3SP was about 3.3% higher than that with the PP in condition 1 and 3.8% in condition 2.

Concerning the overall effort that the listeners allocated during the speech intelligibility task, there was a significant reduction in overall pupil dilation with the P3SP and P3 relative to the PP, suggesting a reduced effort in performing the task when listening with the devices with a higher MFO and a multichannel MFO algorithm. However, some differences in this behavior were observed between condition 1, where the speech level was higher, and, thus, saturation artifacts were more pronounced, and condition 2, where fewer saturation artifacts were present and audibility limited performance. In the following sections, similarities and differences across the two conditions are evaluated and discussed in an attempt to reach a more comprehensive understanding of the effect of higher MFO (P3SP versus P3), increased number of MFO frequency bands (main difference between P3 versus PP[†]), and the combination of both factors (P3SP versus PP) on listening effort.

Peak Pupil Dilation

Because task-evoked pupil dilation has a delay that is normally comprised between 0.5 and 1.5 seconds (Hoeks & Lev-elt 1993), the PPD, which occurred here around 0.5 seconds after sentence offset, can be considered as an indicator of the amount of cognitive resources allocated for the perception of the acoustic signal (Kahneman & Beatty 1966; Beatty 1982; Zekveld et al. 2010). In condition 1, the PPD obtained with the P3SP was significantly lower than the one of the PP, suggesting that already an improved sound quality of the acoustic signal led to a decrease in listening effort. On the contrary, no differences in PPD were observed between P3SP and P3, and between P3 and PP. Hence, the analysis on the PPD suggests that it was the combination of both higher MFO and increased number of MFO frequency bands to reduce listening effort with the P3SP.

The presence of a distorted speech signal with the PP may have degraded the natural ability of exploiting low-level acoustic details for predicting upcoming speech sounds (Winn 2016). Generally, listeners can exploit differences at the syllabic level to predict the upcoming syllable for rapid lexical retrieval (Rön-berg et al. 2013). This bottom-up predicting processing strategy has been shown to improve speech processing in both adults and children (Gow 2002; Mahr et al. 2015). It is possible that the degradation in signal quality due to saturation artifacts with the PP disrupted this predictive processing ability, which, in turn, increased the need to allocate more resources for processing the speech signal. In contrast to condition 1, no differences in PPD were observed in condition 2, suggesting that the cognitive resources utilized during listening did not differ across processors. These results are consistent with the initial hypothesis that decreasing the speech level in condition 2 would reduce the saturation artifacts in the PP. Hence, in contrast to condition 1, similar resources were allocated in condition 2 to process the acoustic signal of similar sound quality with the three sound processors.

Overall Effort for Listening and Preparing the Response (0–5 seconds After Sentence Onset)

Although the PPD is a common parameter to describe pupil dilation, it only captures a single moment in the pupillary response that does not reflect the morphology of the pupillary dilation. Several studies have shown that time-dependent differences in the pupillary response can be detected by modeling the time course of the pupil response via a GCA (Kuchinsky et al. 2013; Winn et al. 2015; Wendt et al. 2018). The GCA was, indeed, shown to reveal time-dependent effects that were not necessarily reflected in the PPD (Wendt et al. 2018).

In the current study, the GCA (0–5 seconds after sentence onset; Fig. 6A; Table 2; Tables 1 and 2 in Supplemental Digital Content 2, <http://links.lww.com/EANDH/A507>) showed that both the P3SP and the P3 led to an overall reduced pupil dilation relative to the PP over the whole time course of the task, which, first, involved listening to the sentence and, later, preparing for the response. The P3SP also showed an overall reduced dilation relative to the P3, suggesting that solely the effect of a higher MFO could decrease the cognitive resources allocated for both listening and response preparation. The overall reduction in pupil dilation with the P3SP versus PP, and P3 versus PP occurred both in condition 1 and in condition 2, while the reduced dilation with the P3SP versus P3 occurred only in condition 1. Hence, the GCA analysis revealed how the cognitive resources allocated for processing speech could be reduced, in condition 1, by a device with solely a higher MFO, and, in condition 2, by the combination of both higher MFO and increased number of MFO frequency bands.

Effort During Response Preparation (2–5 seconds After Sentence Onset)

The response-preparation window, that is, the 3 seconds time window after the PPD and before the response window, reflected the more cognitive dynamics of sentence rehearsal and/or sentence reconstruction (Winn 2016). In this time window, the listeners needed to store the sentence in working memory and, eventually, reconstruct those portions of the sentence that were either not heard or distorted. The Ease of Language Understanding Model (ELU; Rön-berg et al. 2008, 2013) proposes a framework to explain this mechanism. When there is mismatch between the input signal and the long-term memory, as it may be the case in the presence of speech with artifacts, distortions, or inaudible segments, a longer time is required to process the signal. This is referred to as explicit processing and may include different processes, for example, inference-making, semantic integration, storing of information, and inhibition of irrelevant information. These explicit processes typically operate in a time scale of the order of seconds (Rön-berg et al. 2008). If occurring they should, therefore, be reflected in the 3-second window of response preparation.

In both conditions 1 and 2, the GCA in this time window revealed a significantly smaller mean dilation with the P3SP and P3 than the one with the PP. In condition 1, but not in condition 2, the mean dilation with the P3SP was also significantly lower than the one with the P3. Hence, the cognitive resources allocated for rehearsing/reconstructing the sentence were always lower with the P3SP and P3 relative to the PP. This finding suggests that both the higher MFO and increased number of frequency channels helped in reducing the cognitive resources for

[†] P3 has also a slightly higher MFO than the PP, but the difference in signal quality between the two processors is dominated by the increased number of MFO frequency bands in the P3.

sentence retention in working memory. It should be noted that the overall effort allocated in the response-preparation window could be, at least partly, dependent on the listening effort allocated in the preceding listening window.

Focusing the analysis on the response-preparation window allowed exploration of the amount of cognitive resources that hearing-impaired listeners allocated for reconstructing speech. One can assume that few resources are allocated for sentence reconstruction if the pupil size decreases during response preparation, indicating a release of effort between listening to the sentence and the vocal response. Effort release was obtained with all three processors in condition 1, and with the P3 and P3SP in condition 2, as indicated by the significantly negative slopes (Table 3). However, no effort release was obtained with the PP in condition 2, as suggested by a slope of decay that did not differ from zero. After the PPD, the pupil did not decrease in size with the PP, but rather showed a sustained pattern of dilation that stretched toward the response window. A similar pattern of sustained pupil dilation during response preparation was previously observed after a speech signal that was degraded either in spectral resolution (Winn et al. 2015) or in semantic context (Winn 2016). Winn et al. (2015) investigated the effect of degrading the spectral resolution of sentences, via a noise-channel vocoder, on listening effort. Interestingly, the pattern of sustained pupil dilation was only obtained in the condition with the lowest number of vocoder channels, that is, in the condition where speech was degraded the most. Similarly, Winn et al. (2016) showed a pattern of sustained pupil dilation during response preparation in the presence of low-context sentences, suggesting that the lack of semantic context could disrupt the predictive processing of speech and have long-lasting effects on cognitive effort.

Effort Release for Correct Versus Incorrect Responses

To further understand the underlying mechanisms leading to a sustained response after peak dilation, the pupil dilations for correct and incorrect sentences were considered separately (Fig. 7). Following a correct sentence, there was a “pick and release” pattern of dilation, where all three sound processors showed a decay toward baseline in both conditions. Previous studies (Bradshaw 1968; Ahern and Beatty 1979; Zekveld et al. 2010; Winn et al. 2015; Winn 2016) have also observed a “peak and release” pattern of dilation in situations where the listener perceived to have solved a problem, had correctly heard a sentence, had high lexical context, or was more proficient in a language.

On the contrary, following an incorrect sentence, there was a “peak and sustain” pattern of dilation, where the release of effort was either delayed and only occurring in the response window (condition 1), or it was not observed at all (PP in condition 2). In condition 1, the release of effort eventually occurred in the response window for all three processors. A possible interpretation is that the listeners, once they had started to vocalize the response, perceived the task as being “resolved” and did not put further resources for sentence reconstruction, although their final response contained one or more incorrect words. In condition 2, instead, the listeners continued to allocate resources to reconstruct the sentence as if they perceived the task as being “unresolved” throughout both the response preparation and the vocalization of the sentence.

Audibility Limitation

Considering that fewer saturation artifacts occurred in condition 2, other limitations than saturation artifacts had to come into play and lead to a “peak and sustain” response, with the extreme pattern obtained for the PP of a monotonically increasing pupil dilation from the baseline to the response window. The existence of a significant correlation between speech intelligibility performance and PTA (Fig. 4), as well as SL (Fig. 1 Supplemental Digital Content 1, <http://links.lww.com/EANDH/A506>), suggests that the listeners’ performance was limited by audibility with all three sound processors in condition 2. Thus, decreasing the speech level by 5 dB to reduce saturation artifacts in condition 2 caused some segments of speech to be below the audibility threshold. Hence, the listener continued to allocate resources during both the response preparation window and the response window to reconstruct the missing words. This long-lasting allocation of resources suggests that the listener did not have the perception of succeeding in reconstructing the sentence correctly and kept allocating resources for reconstruction even during the vocal response (Bradshaw 1968; Ahern and Beatty 1979; Zekveld et al. 2010; Winn et al. 2015; Winn 2016).

Considering that audibility-related limitations affected the performance in condition 2, one should primarily focus on condition 1 to understand how the MFO of bone-anchored devices affects listening effort and speech intelligibility. Condition 2, which probably reflected both audibility limitations and MFO-related artifacts, further highlights the issue of the limited dynamic range available to BAHS user (Zwartenkot et al. 2014).

Ecological Validity of This Study

The speech levels used in this study were optimally adjusted to obtain saturation artifacts with the PP but not with the P3SP. Due to the relatively low MFO in BAHS, the average speech levels obtained to reach saturation (76 dB SPL in condition 1 and 71 dB SPL in condition 2) corresponded to a loud speech signal but not shouted speech (Pearsons et al. 1977). It has been shown that people with hearing loss are exposed to, on average, a sound pressure level of about 68 dB SPL during a conversation in noise (Wagener et al. 2008). The 75th percentile of the distribution of short-term SPLs for conversations in noise was of about 78 dB SPL. Thus, the average speech levels presented in the current study are within the 75th percentile of the distribution of SPLs that people who are hard of hearing are often exposed to while having a conversation in noise (Wagener et al. 2008). Although this study was specifically designed to investigate the effect of saturation artifacts on listening effort, the utilized speech levels can still be considered ecologically valid.

CONCLUSION

By comparing listening effort during a speech intelligibility task with three BAHS sound processors with a different MFO, this study provides evidence that people with hearing loss allocate less cognitive resources when wearing a device with a higher MFO. These findings demonstrate that most listeners could benefit from the improved sound quality delivered by a device with higher maximum output, in particular in noisy sound environments, both in terms of improved performance and reduced effort, independent of their PTA.

ACKNOWLEDGMENTS

The authors thank Matthew Winn for the helpful discussion of the results.

This work was supported by Oticon Medical AB (Askim, Sweden).

T. R., M. H., T. L., D. W., C. W., and P. M. designed the experiments; C. W. performed the experiments; F. B. and D. W. analyzed the data; F. B. wrote the article and provided the Supplementary Digital Content; all authors discussed the results and provided critical revision on the manuscript.

The authors have no conflicts of interest to disclose.

Address for correspondence: Federica Bianchi, Oticon Medical AB, Kongebakken 9, Smørum, Denmark. E-mail: fedb@oticonmedical.com

Received August 3, 2018; accepted December 18, 2018.

REFERENCES

- Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, *205*, 1289–1292.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol Bull*, *91*, 276–292.
- Bradshaw, J. L. (1968). Pupil size and problem solving. *Q J Exp Psychol*, *20*, 116–122.
- Byrne, D., Dillon, H., Tran, K., et al. (1994). An international comparison of long-term average speech spectra. *J Acoust Soc Am*, *96*, 2108–2120.
- Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear Hear*, *35*, 600–610.
- Downs, D. W. (1982). Effects of hearing and use on speech discrimination and listening effort. *J Speech Hear Disord*, *47*, 189–193.
- Dreschler, W. A. (1988). The effect of specific compression settings on phoneme identification in hearing-impaired subjects. *Scand Audiol*, *17*, 35–43.
- Gow, D. W. J. (2002). Does English coronal place assimilation create lexical ambiguity? *J Exp Psychol Hum Percept Perform* *28*, 163–179.
- Hick, C. B., & Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *J Speech Lang Hear Res*, *45*, 573–584.
- Hodgetts, W. E., & Scollie, S. D. (2017). DSL prescriptive targets for bone conduction devices: Adaptation and comparison to clinical fittings. *Int J Audiol*, *56*, 521–530.
- Hoeks, B., & Levelt, W. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behav Res Methods Instrum Comput*, *25*, 16–26.
- Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hear*, *34*, 523–534.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585.
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Jr, et al. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*, 23–34.
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *J Stat Softw* *82*, 1–26.
- Lunner, T., Hellgren, J., Arlinger, S., et al. (1997). A digital filterbank hearing aid: Three digital signal processing algorithms—user preference and performance. *Ear Hear*, *18*, 373–387.
- Lunner, T., Rudner, M., Rosenbom, T., et al. (2016). Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. *Ear Hear*, *37 Suppl 1*, 145S–154S.
- Mahr, T., McMillan, B. T., Saffran, J. R., et al. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, *142*, 345–350.
- Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. New York, NY: CRC Press.
- Mirman, D., Dixon, J. A., Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: growth curves and individual differences. *J Mem Lang*, *59*, 475–494.
- Nielsen, J. B., & Dau, T. (2011). The Danish hearing in noise test. *Int J Audiol*, *50*, 202–208.
- Ohlenforst, B., Wendt, D., Kramer, S. E., et al. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hear Res*, *365*, 90–99.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., et al. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear Res*, *351*, 68–79.
- Pearsons, K., Bennett, R., Fidell, S. (1977). *Speech Levels in Various Noise Environments*. Washington, D.C.: U.S. Environmental Protection Agency, EPA/600/601–677/025 (NTIS PB270053).
- Pichora-Fuller, M. K., & Kramer, S. E. (2016). Eriksholm workshop on hearing impairment and cognitive energy. *Ear Hear*, *37(Suppl 1)*, 1S–4S.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear*, *37(Suppl 1)*, 5S–27S.
- Piquado, T., Isaacowitz, D., Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*, 560–569.
- Reinfeldt, S., Håkansson, B., Taghavi, H., et al. (2015). New developments in bone-conduction hearing implants: A review. *Med Devices (Auckl)*, *8*, 79–93.
- Rönnerberg, J., Lunner, T., Zekveld, A., et al. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front Syst Neurosci*, *7*, 1–17.
- Rönnerberg, J., Rudner, M., Foo, C., et al. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int J Audiol*, *47(Suppl 2)*, S99–105.
- Sarampalis, A., Kalluri, S., Edwards, B., et al. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J Speech Lang Hear Res*, *52*, 1230–1240.
- Scollie, S., Hodgetts, W., Pumford, J. (2018). DSL for bone anchored hearing devices: Prescriptive targets and verification solutions. *Audiology Online*. Retrieved June 25, 2018.
- Seewald, R., Moodie, S., Scollie, S., et al. (2005). The DSL method for pediatric hearing instrument fitting: Historical perspective and current issues. *Trends Amplif*, *9*, 145–157.
- Smets, K., Wolters, F., Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol*, *26*, 183–196.
- Souza, P., Arehart, K., Neher, T. (2015). Working memory and hearing aid processing: Literature findings, future directions, and clinical applications. *Front Psychol*, *6*, 1894.
- Stone, M. A., & Moore, B. C. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *J Acoust Soc Am*, *121*, 1654–1664.
- Wagener, K. C., Hansen, M., Ludvigsen, C. (2008). Recording and classification of the acoustic environment of hearing aid users. *J Am Acad Audiol*, *19*, 348–370.
- Wendt, D., Hietkamp, R. K., Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear Hear*, *38*, 690–700.
- Wendt, D., Koelewijn, T., Książek, P., et al. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hear Res*, *369*, 67–78.
- Wingfield, A. (2016). Evolution of models of working memory and cognitive resources. *Ear Hear*, *37(Suppl 1)*, 35S–43S.
- Winn, M. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends Hear*, *20*, 1–17.
- Winn, M. B., Edwards, J. R., Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear*, *36*, e153–e165.
- Wu, Y. H., Stangl, E., Chipara, O., et al. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear Hear*, *39*, 293–304.
- Zekveld, A. A., Kramer, S. E., Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear Hear*, *31*, 480–490.
- Zekveld, A. A., Kramer, S. E., Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear Hear*, *32*, 498–510.
- Zwartenkot, J. W., Snik, A. F., Mylanus, E. A., et al. (2014). Amplification options for patients with mixed hearing loss. *Otol Neurotol*, *35*, 221–226.