Perceptual references for independent dimensions of speech quality as measured by electroencephalography

Stefan Uhrig¹ · Sebastian Arndt² · Sebastian Möller¹ · Jan-Niklas Voigt-Antons¹

Received: 10 March 2017 / Published online: 20 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract Quality of transmitted speech can be decomposed into three perceptual dimensions: Noisiness, coloration and discontinuity (Wältermann, Dimension-based quality modeling of transmitted speech. Springer, Berlin. doi:[10.1007/978-3-642-35019-1](https://doi.org/10.1007/978-3-642-35019-1), 2013). The purpose of the present study was to explore whether degradation of speech quality on each perceptual dimension affected the morphological and temporal characteristics of the P300 event-related brain potential (ERP) component. The P300 is composed of two subcomponents, P3a and P3b, which served as neurophysiological indicators of distinct processes in human quality perception (Polich, Neuropsychology of P300. Oxford University Press, Oxford. doi:[10.1093/oxfordhb/9780195374148.013.0089](https://doi.org/10.1093/oxfordhb/9780195374148.013.0089), 2012; Raake and Egger, Quality and quality of experience. Springer International Publishing, Cham, pp. 11–33, 2014): While the earlier P3a reflects attentional processing after the occurrence of novel sensory events, the later P3b is associated with memory operations following the detection of task-relevant stimuli. Electroencephalography (EEG) was used to record the electrical brain activity of subjects

($N = 24$) performing a three-stimulus oddball task. Degraded stimuli were generated from the audio recording of a spoken word. The analysis of P3a- and P3b-related activity at electrode positions Fz, Cz and Pz provided support for the existence of different perceptual references for quality-impaired vs. high-quality stimulus contexts as well as quality degradations on single perceptual dimensions.

Keywords Speech perception · Speech quality · Quality reference · Physiological quality assessment · Electroencephalography (EEG) · P300 · P3a · P3b

Introduction

In telecommunications, quality of transmitted speech is commonly assessed by behavioral measures like the subjective evaluation of stimuli on category rating scales [36]. Such quality ratings are relative in nature, i.e., they reflect a person's judgment about the quality of a test stimulus either in comparison to another presented stimulus [e.g., in the degradation category rating (DCR) or comparison category rating (CCR) paradigms] or to a long-standing internal reference [e.g., when being requested in an absolute category rating (ACR) paradigm] [25]. Although these measures reliably capture the behavioral descriptions of quality judgments, they cannot be utilized to study the mostly unconscious processes involved in the formation of these judgments within the experiencing person [37, 38].

A new approach in quality assessment tries to overcome this limitation by employing non-invasive electrophysiological methods to measure changes in brain activity associated with internal quality-related representations and processes [2, 20]. Electroencephalography (EEG) proved to

✉ Stefan Uhrig
stefan.uhrig@qu.tu-berlin.de

Sebastian Arndt
sebastian.arndt@ntnu.no

Sebastian Möller
sebastian.moeller@tu-berlin.de

Jan-Niklas Voigt-Antons
jan-niklas.voigt-antons@tu-berlin.de

¹ Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany

² Norwegian University of Science and Technology, Trondheim, Norway

be the most frequently employed method due to its high temporal resolution, showing modulations of specific frequency bands in the continuous EEG as well as effects on the morphological and temporal characteristics of event-related brain potential (ERP) components [21, 47].

A number of prior studies have demonstrated effects of stimulus quality on brain activity. In the auditory modality, long-term exposure to degraded speech stimuli was associated with an increase in the relative power of alpha and theta frequency bands located in frontal and parietal-occipital areas, reflecting the influence of emotional processing and mental fatigue [3, 5, 8]. Visual perception was shown to be affected by the perceived quality of images and videos [12, 13, 33, 34, 39, 40, 58] on both explicit and implicit visual processing stages [41], also exerting an influence on the person's affective state [32]. Moreover, first attempts have been made to investigate the neural correlates of audiovisual quality [7, 9–11].

Rapid, short-term changes in neuro-electrical activity time-locked to the onset of a defined sensory, cognitive, affective or motor event manifest in the averaged EEG signal as a stereotypical sequence of ERP components [35, 71]. Specifically the P300 component refers to a positive voltage change occurring approx. 300 ms after the unexpected onset of a meaningful event, which is presumably caused by increased neuro-inhibitory activity [49–51, 54, 64]. The P300 waveform results from the superposition of the waveforms of two subcomponents P3a and P3b, both of which typically show middle latencies, with the P3a having a shorter latency than the P3b. While the P3a reflects the allocation of attentional resources—indicating the strength and speed of the involuntary orienting response elicited by novel stimuli [23, 42, 61]—the P3b has been associated with stimulus categorization and memory operations. Accordingly, distinct neural sources associated with attention and memory functions were shown to be involved in the generation of the P3a and P3b, respectively [24, 62, 66, 72].

Usually the P300 or its subcomponents are evoked by applying a variant of the classic oddball paradigm. The three-stimulus oddball paradigm is based on the sequential presentation of stimuli, each of which belongs to one of three different stimulus types: A frequent, task-irrelevant standard (S), an infrequent, task-relevant target (T) and an infrequent, task-irrelevant distractor (D). Due to their unexpected occurrence, T and D are jointly referred to as “oddballs”. Subjects are instructed to respond only to the occurrence of T. Presentation of T triggers the P300 (i.e., the superposition of the P3a and P3b), presentation of D triggers only the P3a. Since the P300 constitutes a very robust and well-studied neuro-electrical phenomenon, an extensive literature base including methodological standards is readily available (e.g., see [17] for the clinical domain).

Using speech stimuli of different complexity (phonemes, words, sentences), previous studies revealed positive and negative effects of degradation intensity (e.g., discrete levels of signal-to-noise ratio, SNR) on P300 amplitude and latency [4, 6, 56]. Interestingly, stimuli with only a small degree of quality degradation, which were explicitly judged as having normal quality, have been shown to affect the P300 characteristics in the same way as highly-impaired stimuli, which were judged as degraded. This implies that the degree of quality degradation has to rise above a certain threshold in order to be perceived (and judged) consciously, yet already influences early implicit processes and probably favors the generation of discomfort, stress and fatigue [4, 55]. So far, cumulative evidence suggests that the P300 can reliably indicate differences in quality percepts and judgments, but it still remains an open question whether its subcomponents P3a and P3b are equally suitable as indicators of quality-related processes.

Quality degradations of speech signals can be of different perceptual nature, e.g., a speech signal may be impaired by additive noise, frequency distortions, interruptions, or alike. Depending on the strength of the degradation, all these impairments may result in the same quality judgment (e.g., a moderately degraded speech signal, obtaining a “fair” rating on a five-point category scale according to ITU-R Recommendation BS.1770-3 [25]), although being perceptually quite different: Additive noise will result in a perception of “noisiness”, frequency distortions in a perception of “coloration”, and interruptions in a perception of “discontinuity” [69, 70]. To analyze this observation, multidimensional analyses have been carried out based on multidimensional scaling or semantic differential paradigms, revealing three [68–70] or five [59, 60] independent dimensions. Such perceptual dimensions can be utilized to diagnose the sources of impairments, thus providing help for service optimization. However, it still remains unclear how perceptual dimensions of quality are reflected in the human brain, i.e., whether they provoke the same type and strength of neural responses.

The present study aimed at exploring the effects of speech quality on the neural response characteristics of the P300 subcomponents, P3a and P3b. In particular, it focused on answering the following joint questions: First, based on the assumption that the three independent dimensions noisiness, coloration and discontinuity underly the quality perception of transmitted speech signals and therefore have to be regarded as separate perceptual quality features [68]: Are these three dimensions represented as different perceptual references within the experiencing person? And second, assuming that both P3a and P3b indicate attentional and memory-related processes involved in the formation of quality percepts: Does manipulation of the stimulus environment in terms of quality or a single-

dimensional degradation of the perceptual reference modulate the morphological and temporal characteristics of both subcomponents?

A simplified model of quality perception sketched in Fig. 1 provided the theoretical framework for the present study, combining a model of internal quality formation [57] with a model of cognitive P300 activity [50, 51]. As the three-stimulus oddball paradigm requires the perceptual discrimination of stimuli, triggering of the P3a and P3b was anticipated during the early perception stage of the quality formation process. After *sensory processing & perception* of a physical event present within a certain contextual setting, the resulting perceptual event is compared with the perceptual reference, i.e., an internal representation of the current environment (“stimulus context”), containing all recently perceived stimuli (*anticipation & matching*). In case an oddball (T, D) was presented, the perceptual event deviates from the perceptual reference and attentional resources are allocated for the upcoming orienting response, which evokes the P3a. If the oddball is further classified as task-relevant (T), subsequent *memory processing* causes the updating and storage of the perceptual reference, eliciting the P3b. Finally, *response-related processing* leads to the generation of a task-appropriate motor response.

Methods

The design of the present study followed practical guidelines concerning the standardized application of the EEG method to quality assessment [1, 2]. These guidelines recommend the use of minimal stimulus sets, which include only a small number of stimulus types with short presentation time and synchronous onset, as well as the employment of well-established experimental paradigms with sufficiently high reliability like the classic oddball paradigm [67]. Data collection was preceded by a short pretest on one member of the laboratory to demonstrate the effectiveness of the planned manipulations and to refine the experimental setup [e.g., intensity and duration of stimuli, duration of the inter-stimulus interval (ISI), number of trials per block].

Subjects

24 subjects participated in the study (11 males, 13 females; age: $M = 28.96$ years, $SD = 6.69$, $R = 29$; 87.5% right-handed), all of whom were native German speakers with no self-reported hearing problems. All subjects gave their informed consent after reading a short information about the EEG method provided by the Deutsche Gesellschaft für Psychologie (DGPs; English: German Psychological

Society) and received a monetary compensation at the end of the experiment.

Stimuli

A clean audio recording of the German word /Haus/ (English: “house”) spoken by a female voice in a sound-insulated environment with a high-quality microphone (AKG Acoustics, C414 B-XLS/ST) was chosen as stimulus material for the experiment due to its neutral meaning as well as clear and strong onset (duration: 500 ms, sampling rate: 16 kHz, bit rate: 16 bps).

Three quality-impaired stimuli were produced from the audio speech file by degrading only one of the three perceptual dimensions (noisiness, coloration, discontinuity) at a time:

1. Signal-correlated noise generated according to ITU-T Recommendation P.810 [27] with a distortion rate of 18 Q resulted in a noisiness-impaired stimulus (N).
2. Application of a bandpass filter with a low-cutoff of 1 kHz and a high-cutoff of 2 kHz resulted in a coloration-impaired stimulus (C).
3. Packet loss generated according to ITU-T Recommendation G.722.2 [26] with a packet loss rate of 20% resulted in a discontinuity-impaired stimulus (P).

Besides, the original recording served as a high-quality stimulus (HQ). The intensities of all stimuli were normalized according to ITU-R Recommendation BS.1770-3 [25], resulting in a target loudness of -23 LUFS (Loudness Units relative to Full Scale).

It was assumed that the selected degradation methods (1.–3.) would create a noticeable change in the perceived quality on each perceptual dimension. This quality manipulation was checked empirically by using standardized category rating paradigms (see Sect. 2.4). In addition, an instrumental estimation of speech quality based on the perceptual objective listening quality assessment (POLQA) model was conducted for each stimulus in accordance with ITU-T Recommendation P.863 [28]. The application of the POLQA model (3 repetitions, duration: 1.5 s) resulted in predicted MOS values for HQ (4.21), N (1.92), C (3.95) and P (1.20), which indicate relevant differences in the predicted speech quality of the four stimuli.

Experimental design and hypotheses

The present study adopted the three-stimulus oddball paradigm to investigate the impact of independent perceptual dimensions of speech quality on P3a and P3b characteristics. It was assumed that the abrupt presentation of an oddball (T, D) would cause an orienting response, which in case of T would be succeeded by memory

Fig. 1 Simplified model of quality perception (*circles* internal processes; *parallel horizontal lines* internal representations; *rectangles* input signals, output), based on a conceptual model of internal quality formation [57] and a model of cognitive P300 activity [51]

operations. After T, both P3a and P3b waveforms should be triggered, producing the superposed P300 waveform. After D, only the P3a waveform should be evoked.

Six stimulus sets resulted from the combination of the degraded stimuli (N, C, P) and stimulus types (S, T, D) without repetition. Six additional stimulus sets were generated with HQ as S, combining the degraded stimuli and oddballs (T, D) without repetition. No stimulus sets with HQ as oddball were used, since further increasing the number of stimulus sets would result in exhaustively long test sessions with a higher probability of interference by mental fatigue and drowsiness. Table 1 lists all stimulus sets used in the present study. In the following sections, particular stimulus sets are named according to their stimulus-to-stimulus-type assignments in the order of S-T-D, e.g., N-C-P being the stimulus set with N as S, C as T and P as D, respectively.

Hypotheses about the effects on P3a and P3b characteristics were derived from the simplified model of quality perception depicted in Fig. 1. According to this model, the repeated presentation of degraded or HQ stimuli serving as S should lead to the formation of different perceptual references, i.e., different internal representations for each degraded or HQ stimulus. Hence, the deviation of a certain degraded stimulus serving as oddball from the constant perceptual reference (*anticipation & matching*) varies depending on which dimension of the quality percept has

changed. It was presumed that the quality of the stimulus context would affect the morphological and temporal characteristics of the P3a. Following the presentation of T, memory processes responsible for the updating of the perceptual reference (*memory processing*) should vary depending on which represented dimension they operate on. Therefore, effects of the perceptual references on P3b characteristics were expected. Due to the exploratory nature of the study, the directions of the anticipated effects on both P3a and P3b characteristics remained unspecified.

Experimental procedure

Each experimental session started with an interview, during which subjects were asked to report their handedness, potential hearing problems, vision correction, drug consumption and hours of sleep they had last night.

During the oddball task, word stimuli were presented in a pseudo-randomized order such that at least one S fell between two consecutive oddballs, and subjects responded to the occurrence of T by pressing a button box. Moreover, two category rating paradigms were performed in accordance with ITU-R Recommendation BS.1770-3 [25]: an absolute category rating (ACR) task and a comparison category rating (CCR) task—one before and the other one after the oddball task, in randomized order across subjects. During the ACR task, a test stimulus was presented, whose quality had to be evaluated on a unipolar discrete five-point rating scale (mean opinion score, MOS) with the values “1”, “2”, “3”, “4” and “5” as well as respective category labels in German “schlecht”, “dürftig”, “ordentlich”, “gut” and “ausgezeichnet” (English: “bad”, “poor”, “fair”, “good” and “excellent”). Each of the four stimuli used in the oddball task (HQ, N, C, P) occurred five times, stimulus presentation was carried out in randomized order. During the CCR task, a reference stimulus and a test stimulus were presented sequentially with a short pause of 1 s in between, afterwards subjects had to evaluate the quality of the second stimulus relative to the quality of the first stimulus on a bipolar discrete seven-point rating scale (comparison mean opinion score, CMOS) with the values “−3”, “−2”, “−1”, “0”, “1”, “2” and “3” as well as corresponding labels in German “viel schlechter”, “schlechter”, “etwas schlechter”, “etwa gleich”, “etwas besser”, “besser” and “viel besser” (English: “much worse”, “worse”, “slightly worse”, “about the same”, “slightly better”, “better” and “much better”). The 16 pairings of stimuli—including pairs of the same stimulus (HQ-HQ, N-N, C-C, P-P)—were presented twice, each time in randomized order. The subjects delivered their quality judgments by pressing buttons on a keyboard, which were labeled with the values of MOS and CMOS, respectively. Subfigure 1 in Fig. 2 illustrates the order of tasks to be performed by each subject.

Table 1 Twelve stimulus sets used in the experiment

	Stimulus type				<i>n</i>
	S	T	D		
Stimulus set	1	N	C	P	300
	2	N	P	C	300
	3	C	N	P	300
	4	C	P	N	300
	5	P	N	C	300
	6	P	C	N	300
	7	HQ	C	P	100
	8	HQ	P	C	100
	9	HQ	N	P	100
	10	HQ	P	N	100
	11	HQ	N	C	100
	12	HQ	C	N	100

The first six stimulus sets represent combinations of the degraded stimuli (N noisiness-impaired, C coloration-impaired, P discontinuity-impaired) and stimulus types (S standard, T target, D distractor). The other six stimulus sets consist of the high-quality stimulus (HQ) as S and combinations of the degraded stimuli and “oddball” stimulus types (T, D). The last column denotes the number of trials per stimulus set (*n*)

Fig. 2 1 Experimental procedure consisting of three parts: the first part, category rating 1, being either an absolute category rating (ACR) or comparison category rating (CCR) task. The second part, oddball, being a three-stimulus oddball task consisting of twelve consecutive blocks with changing stimulus-to-stimulus-type assignments. The third part, category rating 2, being either a CCR or ACR task, depending on which category rating task was used in category rating 1. 2 Visual illustration of the three-stimulus oddball task. Quality-impaired stimuli generated from an audio recording of the spoken word /Haus/ are presented in a randomized sequence. Each stimulus (*black*, *green*, *red*) is assigned to one of three stimulus types, standard (S), target (T) and distractor (D), which vary in probability and task relevance. Subjects are instructed to respond only to the occurrence of T. 3 Schematic ERP waveforms of the P3a and P3b subcomponents, whose superposition produces the P300 waveform. Presentation of S (*dotted line*) triggers no P300-related activity, presentation of T (*solid line*) triggers both P3a and P3b, presentation of D (*dashed line*) triggers only P3a. Subfigures 2 and 3 are based on Figure 7.1 (Subfigure C) in a review on the P300 literature [51]

At the beginning of the oddball task, subjects were instructed to press the button box as fast and as accurately as possible whenever they detected T. Within each block only stimulus-to-stimulus-type assignments of a particular stimulus set were presented. Prior to the start of a new block, subjects were introduced to the new stimulus-to-stimulus-type assignments: First by passively listening to the repetitions of each stimulus type, then by actively performing a sequence of demo trials, after which they

could start the test trials. A total number of 300 trials per block was presented with a 70:20:10 frequency ratio for the three stimulus types S, T and D. Each trial consisted of the presentation of a physical event, followed by an inter-stimulus interval (ISI) of 1200 ms plus/minus a jitter of 200 ms. Sequences of experimental blocks (stimulus sets) and trials were randomized for each session. In total, subjects ran through 12 blocks: Six stimulus sets with a degraded stimulus as S, each consisting of 300 trials, and six

stimulus sets with HQ as S, each consisting of only 100 trials to reduce the total length of the experiment. Subfigure 2 in Fig. 2 illustrates a randomized sequence of stimulus presentations during the oddball task with Subfigure 3 depicting the composition of the associated neural responses.

All test sessions were carried out in a quiet standardized test room in accordance with ITU-R Recommendation BS.1770-3 [25]. During an entire session subjects sat in front of a monitor, wearing an EEG electrode cap as well as headphones (Sennheiser HD 280) through which stimuli were presented binaurally. The subjects themselves were allowed to set their preferred loudness level at the beginning of the first category rating task in order to create a more ecologically valid listening situation. During each block subjects held the button box in their dominant hand and kept their eyes open. The monitor screen was used to display instructions for all tasks as well as requesting a response from the subject after hearing the test stimulus (ACR) or stimulus pair (CCR) and otherwise turned black. After each block subjects were offered to take a short break. Including the time to prepare the equipment and conduct the whole experiment, a single test session took about two and a half hours.

Data collection

Running on a standard laboratory PC, Presentation® software (version 18.3; Neurobehavioral Systems, Inc., USA) executed stimulus presentation, sending event triggers to the EEG amplifier. Electrocardiac activity was measured at 16 scalp positions (FP1, F3, C3, P3, TP9, FP2, F4, C4, P4, TP10, Fz, FCz, Cz, CPz, Pz, Oz) according to the extended 10–20 system [44] via a medical-grade EEG recording system (g.GAMMASys; g.tec medical engineering GmbH, Austria) with active electrodes (g.LADYbird). The reference electrode was placed on the left mastoid, the ground electrode on the right mastoid. The sampling rate was set to 512 Hz.

Data processing and analysis

The recorded EEG data was processed by using the EEGLAB toolbox (v13.6.5b) for MATLAB software (R2015b).

First, the raw signal was down-sampled from 512 to 256 Hz, then filtered with a finite impulse response (FIR) filter—setting the low-cutoff to 0.1 Hz and the high-cutoff to 40 Hz—as well as re-referenced to average reference. The extracted epochs were stimulus-locked with a pre-stimulus time of 200 ms serving as baseline activity and a post-stimulus time of 1000 ms. Maximum and minimum thresholds for the signal amplitude were specified for each subject across all channels, such that at most 10% of

epochs would be rejected. Epochs exceeding this threshold were considered to be artifacts and rejected from further analysis.

Ocular artifacts caused by eye movements and blinks were corrected individually for each subject: after computing an independent component analysis (ICA), components associated with ocular activity were subtracted from the data.

Single-subject averaging of single-trial epochs resulted in single-subject average waveforms for each subject, channel, stimulus set and stimulus type. Subsequent grand averaging of single-trial epochs resulted in grand average waveforms for each channel, stimulus set and stimulus type. Average difference epochs T-S and D-S were computed via the following processing steps: Because the number of epochs with S as sensory event (“S epochs”) was always exceeding the number of epochs with T or D as sensory event (“oddball epochs”), both numbers had to be equalized first. This was achieved by randomly selecting S epochs of the same number as oddball epochs. Subtraction of single-trial S epochs from single-trial oddball epochs resulted in single-trial difference epochs, across which the average was taken and variance measures—needed for computing confidence bands and multiple paired t-tests across the time course of the difference epochs (“running t-tests”)—were calculated.

The anticipated effects on P3a and P3b characteristics explicated in Sect. 2.3 could be examined by comparing stimulus sets with degraded stimuli as S and corresponding stimulus sets with HQ as S, given a constant ERP difference and oddball stimulus. Two common ERP parameters, peak amplitude and peak latency, were chosen to quantify the morphological and temporal characteristics of the P3a and P3b subcomponents, respectively. Peak amplitude is defined as the voltage difference between a certain positive/negative maximum (“peak”) in post-stimulus time and the (pre-stimulus) baseline, while peak latency refers to the time passing between the onset of the stimulus and a certain peak. A first analysis step aimed at determining whether different perceptual references exist for quality-impaired vs. HQ stimulus contexts by comparing stimulus sets listed in Table 2. Thus, for each ERP difference (T-S/D-S) and oddball stimulus (N/C/P) the effects of *impairment* (Impaired, HQ) on P3a and P3b peak parameters were analyzed. A second analysis step further examined, whether different perceptual references exist for each of the three perceptual dimensions of speech quality via stimulus set comparisons summarized in Table 3. Hence, for each ERP difference (T-S/D-S) and oddball stimulus (N/C/P), the effects of *perceptual reference* (C, P, HQ/N, P, HQ/N, C, HQ) on P3a and P3b peak parameters were analyzed. For each analysis step, stimulus sets with HQ as S were merged prior to computing the difference waveforms as

Table 2 Stimulus set comparisons to examine effects of *impairment* on P3a and P3b characteristics. Summation in brackets symbolizes merging of stimulus sets prior to averaging

Difference	Oddball	Impairment	Set comparison
T-S	N	Impaired vs. HQ	(3 + 5) vs. (9 + 11)
	C	Impaired vs. HQ	(1 + 6) vs. (7 + 12)
	P	Impaired vs. HQ	(2 + 4) vs. (8 + 10)
D-S	N	Impaired vs. HQ	(4 + 6) vs. (10 + 12)
	C	Impaired vs. HQ	(2 + 5) vs. (8 + 11)
	P	Impaired vs. HQ	(1 + 3) vs. (7 + 9)

Table 3 Stimulus set comparisons to examine effects of *perceptual reference* on P3a and P3b characteristics. Summation in brackets symbolizes merging of stimulus sets prior to averaging

Difference	Oddball	Impairment	Set comparison
T-S	N	C vs. P vs. HQ	3 vs. 5 vs. (9 + 11)
	C	N vs. P vs. HQ	1 vs. 6 vs. (7 + 12)
	P	N vs. C vs. HQ	2 vs. 4 vs. (8 + 10)
D-S	N	C vs. P vs. HQ	4 vs. 6 vs. (10 + 12)
	C	N vs. P vs. HQ	2 vs. 5 vs. (8 + 11)
	P	N vs. C vs. HQ	1 vs. 3 vs. (7 + 9)

outlined above. To compare neural responses in quality-impaired vs. HQ stimulus contexts (first analysis step), also those two stimulus sets with degraded stimuli as S were merged beforehand. Separate time windows, from 350 to 550 ms and from 550 to 750 ms, were used for ERP peak parameter extraction in order to quantify post-stimulus activity related to the P3a and P3b subcomponents, respectively. Both time windows were selected through visual inspection of the grand average T-S and D-S difference waveforms as well as based on the significant results obtained from the running t-tests (see Sect. 3.2).

Two response measures, target response time and hit rate, were chosen to quantify the subject's response behavior during the oddball task, the latter being defined as the relative frequency of responses to T, i.e.,

$$\text{hit rate} = \frac{\text{number of responses to T}}{\text{total number of T}}.$$

Response times less than 1 s were extracted for responses to T in each experimental block (stimulus set).

Using RStudio software (version 1.0.143), independent repeated-measures analyses of variance (ANOVA) were computed for behavioral (category ratings, response measures) as well as electrophysiological data (ERP parameters), each based on a significance level of $\alpha = 0.05$:

- To analyze whether the four stimuli differed in absolute or relative quality, two repeated-measures ANOVAs

were calculated with *stimulus* (HQ, N, C, P) as within-subject factor and either MOS (ACR) or CMOS (CCR) as dependent variable.

- To analyze the effects on response measures, two repeated-measures ANOVAs were calculated with either *block* (1–12) or *stimulus set* (see Table 1) as within-subject factor and either target response time or hit rate as dependent variable.
- To analyze the effects on P3a and P3b peak parameters, 48 repeated-measures ANOVAs were carried out with either *perceptual reference* (C, P, HQ/N, P, HQ/N, C, HQ) or *impairment* (Impaired, HQ) as within-subject factor and either peak amplitude or peak latency as dependent variable for each combination of channel (Fz, Cz, Pz; see Sect. 3.2), ERP difference (T-S/D-S), oddball stimulus (N/C/P) and time window (350–550/550–750).

In case of violating the assumption of sphericity, degrees of freedom were corrected according to Huynh-Feldt (ϵ_{HF}). Generalized eta squared (η_{G}^2) was calculated as a measure of effect size [14, 45]. For all post-hoc pairwise comparisons, multiple paired t-tests with Šidák correction were computed.

Results

Behavioral data analysis

The analysis of the category ratings showed significant effects of *stimulus* on MOS ($F[3, 69] = 129.316$, $p < 0.001$, $\eta_{\text{G}}^2 = 0.755$) and CMOS ($F[3, 69] = 127.844$, $p < 0.001$, $\eta_{\text{G}}^2 = 0.846$). Post-hoc comparisons revealed significant differences between all stimuli for both category ratings, except HQ vs. C, which was only significant for MOS ($p = 0.006$). Figure 3 illustrates average MOS and CMOS as a function of *stimulus*.

The analysis of the response measures revealed significant effects of *block* ($F[11, 253] = 2.406$, $\epsilon_{\text{HF}} = 0.708$, $p = 0.018$, $\eta_{\text{G}}^2 = 0.033$) and *stimulus set* ($F[11, 253] = 2.719$, $\epsilon_{\text{HF}} = 0.448$, $p = 0.034$, $\eta_{\text{G}}^2 = 0.037$) on target response time as well as a significant effect of *stimulus set* on hit rate ($F[11, 253] = 6.059$, $\epsilon_{\text{HF}} = 0.334$, $p < 0.001$, $\eta_{\text{G}}^2 = 0.155$). Post-hoc comparisons revealed significant differences in average target response times for blocks 2 vs. 9 ($p = 0.020$), 2 vs. 10 ($p = 0.012$) and 2 vs. 12 ($p = 0.044$) as well as for stimulus sets HQ-P-C vs. N-C-P ($p = 0.010$) and HQ-P-C vs. HQ-N-P ($p = 0.041$). No significant post-hoc effects on hit rate were found. Figure 4 illustrates average target response times and hit rates as a function of *block* and *stimulus set*.

Fig. 3 Effects of *stimulus* on average mean opinion score (MOS) and comparison opinion score (CMOS) as measured in absolute category rating (ACR) and comparison category rating (CCR) tasks, respectively. Error bars represent 95% confidence intervals

Fig. 4 Effects of *block* and *stimulus set* on target response time and hit rate in the oddball task. Error bars represent 95% confidence intervals

ERP plots

Visual inspection of single-subject average ERP plots guided the exclusion of three subjects due to technical artifacts and unacceptably high levels of noise in their recordings.

Grand average ERP plots were generated for three electrode positions Fz, Cz and Pz, at which the topographical P300 characteristics should be most prominent (“P300 scalp distribution”) [29, 52], displaying the average neural responses across the remaining 21 subjects. Subfigure 1 in Fig. 5 shows grand average ERP waveforms

for each stimulus type in stimulus set C-N-P at electrode position Cz to exemplify the successful triggering of the P3a and P3b.

In addition, grand average ERP difference plots were generated to enable the selection of time windows for the extraction of P3a and P3b peak parameters: While the T-S waveform should indicate superposed post-stimulus activity of both P3a and P3b (i.e., the P300), the D-S waveform should indicate only P3a-related activity. Accordingly, subfigure 2 in Fig. 5 shows grand average ERP difference waveforms T-S and D-S for stimulus set C-N-P at electrode

Fig. 5 1 Grand average ERP waveforms S, T and D for stimulus set C-N-P at electrode position Cz, indicating average neural responses to stimulus types S, T and D. 2 Grand average ERP difference waveforms T-S and D-S for stimulus set C-N-P at electrode position Cz, indicating average neural responses to each stimulus type and average deviations of neural responses to T and D from the neural response to S. Prior to plotting, a second artifact rejection with an absolute threshold of 250 was performed on the electrophysiological data. The legend contains the number of trials per waveform. *Colored arrows* point at assumed peaks of P3a and P3b subcomponents in both waveforms. *Colored lines of dots* (above the time axis) indicate, at which time points the ERP difference waveforms significantly deviate from the baseline (based on multiple paired t-tests with Šidák-adjusted $\alpha = 0.05$). *Error bands* represent 95% confidence intervals

position Cz. Running t-tests with Šidák-adjusted $\alpha = 0.05$ were calculated across the epoch time course to indicate, whenever each waveform significantly deviated from baseline.

Visual inspection of the grand average waveforms as well as significant results from the running t-tests enabled the localization of both subcomponents: In the T-S waveform (red waveforms in subfigures 1 and 2), the potential

peaks around 460 ms and 600 ms should reflect P3a- and P3b-related neural activity, respectively. In the D-S waveform, only one potential peaks manifested around 440 ms, which should indicate P3a-related neural activity.

Electrophysiological data analysis

The analysis of the electrophysiological data was based on comparisons of specific stimulus sets (see Tables 2, 3) in order to isolate effects of *impairment* and *perceptual reference* on P3a and P3b peak parameters (peak amplitude, peak latency), which were extracted from two separate time windows (350–550/550–750). Figure 6 illustrates grand average ERP difference waveform plots for each stimulus set comparison listed in Table 3 at electrode position Cz. These plots visualize average neural response differences T-S (1) and D-S (2) for different stimuli serving as S (*perceptual reference*), given a constant degraded stimulus (N, C, P) as T (1) and D (2), respectively.

Table 4 and Fig. 7 show significant results from four independently conducted ANOVAs of the effects of *impairment* on peak amplitude and peak latency ordered by channel, ERP difference, oddball stimulus, time window of ERP quantification as well as the extracted ERP parameter.

Accordingly, Tables 5, 6 as well as Fig. 8 show significant results from eight independently conducted ANOVAs of the effects of *perceptual reference* on peak amplitude and peak latency.

Discussion

The goal of the present study was to explore the neuro-electrical correlates of perceptual references for three independent dimensions of speech quality: Noisiness, coloration and discontinuity [69, 70]. Perceptual references are considered to be part of the early perception stage of the quality formation process [57], exerting an influence on attentional and memory-related processing. The subcomponents of the P300, P3a and P3b, were chosen as neurophysiological indicators of these early quality-related processes (see Fig. 1). An audio recording of the German word |Haus| spoken by a female voice was degraded on each of the three perceptual dimensions at a time, producing three quality-impaired stimuli (N = noisiness-impaired, C = coloration-impaired, P = discontinuity-impaired) with the original recording serving as a high-quality (HQ) stimulus. In order to test the effects of HQ vs. quality-impaired stimulus contexts as well as single perceptual references on P3a and P3b characteristics, 24 subjects were invited to perform an

Fig. 6 Grand average T-S (1) and D-S (2) difference waveforms for each oddball stimulus (N, C, P) at electrode position Cz, indicating average deviations of neural responses to T (1) and D (2) from neural responses to S. Prior to plotting, a second artifact rejection with an absolute threshold of 250 was performed on the electrophysiological

data. The legend contains the number of trials for each waveform. *Colored lines of dots* (above the time axis) indicate, at which time points the ERP difference waveforms significantly deviate from the baseline (based on multiple paired t-tests with Šidák-adjusted $\alpha = 0.05$). *Error bands* represent 95% confidence intervals

Table 4 Effects of *impairment* on ERP parameters (peak amplitude, peak latency)

	Channel	Difference	Oddball	Reference	Extraction	Peak	df_n	df_d	F	p	η_G^2
1	Cz	T-S	N	Impaired vs. HQ	350–550	Amplitude	1	20	5.407	0.031*	0.085
2	Cz	D-S	C	Impaired vs. HQ	350–550	Amplitude	1	20	4.364	0.050	0.124
3	Pz	T-S	P	Impaired vs. HQ	550–750	Latency	1	20	6.092	0.023*	0.123
4	Pz	D-S	N	Impaired vs. HQ	550–750	Latency	1	20	4.705	0.042*	0.095

Bold indices refer to Fig. 7

* indicate statistical significance

active three-stimulus oddball task with the degraded stimuli (N, C, P) being assigned to three stimulus types, standard (S), target (T) and distractor (D). During the

oddball task, each subject's electrical brain activity was recorded via EEG. Statistical analyses of the resulting behavioral (category rating scores, response measures)

Fig. 7 Arithmetic means (M) and standard deviations (SD) of the effects of *impairment* on ERP parameters (peak amplitude, peak latency). Error bars represent 95% confidence intervals

Table 5 Effects of *perceptual reference* on ERP parameters (peak amplitude, peak latency)

	Channel	Difference	Oddball	References	Extraction	Peak	df_n	df_d	ϵ_{HF}	F	p	η_G^2
1	Fz	T-S	N	C vs. P vs. HQ	350–550	Latency	2	40	–	4.933	0.012*	0.118
2	Fz	T-S	C	N vs. P vs. HQ	350–550	Latency	2	40	–	3.466	0.041*	0.068
3	Fz	D-S	C	N vs. P vs. HQ	350–550	Latency	2	40	–	5.160	0.010*	0.130
4	Cz	T-S	N	C vs. P vs. HQ	350–550	Amplitude	2	40	0.743	3.591	0.052	0.093
5	Cz	T-S	C	N vs. P vs. HQ	350–550	Latency	2	40	1.057	4.229	0.022*	0.093
6	Cz	D-S	C	N vs. P vs. HQ	350–550	Amplitude	2	40	0.788	4.623	0.024*	0.148
7	Cz	D-S	C	N vs. P vs. HQ	550–750	Amplitude	2	40	0.697	4.628	0.029*	0.141
8	Pz	D-S	C	N vs. P vs. HQ	550–750	Latency	2	40	–	4.813	0.013*	0.138

Bold indices refer to Table 6 and Fig. 8

* indicate statistical significance

Table 6 Post-hoc comparisons of the effects of *perceptual reference* on ERP parameters (peak amplitude, peak latency)

1	C	P	2	N	P	3	N	P	4	C	P
P	0.180	-	P	0.438	-	P	0.669	-	P	0.290	-
HQ	0.180	0.030*	HQ	0.294	0.294	HQ	0.028*	0.028*	HQ	0.290	0.080
5	N	P	6	N	P	7	N	P	8	N	P
P	0.051	-	P	0.802	-	P	0.682	-	P	0.002*	-
HQ	0.895	0.072	HQ	0.074	0.080	HQ	0.094	0.094	HQ	0.147	0.352

* indicate statistical significance

and electrophysiological data (ERP parameters) revealed several significant effects.

Behavioral effects: category rating scores

At first, the degradation method applied to the original audio recording affected both the mean opinion score (ACR) and comparison mean opinion score (CCR), with P and N differing significantly from HQ and C. Hence, quality percepts of HQ and C might be described

subjectively as “normal” or “close-to-normal”, while quality percepts of N and P appeared to be “clearly degraded”. Both effects demonstrated the successful manipulation of speech quality in the present experiment and were considered in the following analyses of the response measures and the electrophysiological data. Moreover, the observed MOS values matched the predicted MOS values obtained by applying the POLQA model to the four stimuli (see again Sect. 2.2). In general, the observed differences in quality judgments were quite

Fig. 8 Arithmetic means (M) and standard deviations (SD) of the effects of *perceptual reference* on ERP parameters (peak amplitude, peak latency). Error bars represent 95% confidence intervals

robust with only small variability between subjects (see Fig. 3). This might be an indication that the individual adjustment to a preferred loudness level did not significantly influence quality perception and evaluation of the presented stimuli.

Behavioral effects: response measures

During the oddball task, only weak time-on-task effects (probably due to an increase in mental fatigue and/or drowsiness) were observed, manifesting as a general reduction in average target response times over the course of the experiment. In contrast, the target detection performance as quantified by average hit rates remained unaffected by the amount of time spend on the task, which is not surprising, given that the cognitive load required for optimal task performance should be relatively low in healthy subjects. Furthermore, both response measures were shown to depend on the particular stimulus set: Average responses to T in stimulus set HQ-P-C were faster than in N-C-P and HQ-N-P. Since HQ and P were judged as having the highest and lowest quality of all stimuli, respectively, the difficulty to perceptually discriminate between S and T (T-S discriminability) should be minimal in HQ-P-C, reducing the average target response time and increasing the average hit rate. On the other hand, stimulus sets HQ-C-P and HQ-C-N demonstrated slower and more erroneous responses to T compared to the other stimulus sets with HQ as S, which corresponds with the observation

that HQ and C were judged similarly on MOS/CMOS. Besides, variability increased drastically in these more demanding stimulus sets, pointing to relevant differences in auditory discrimination ability between subjects.

Electrophysiological effects: perceptual references for quality-impaired vs. high-quality stimulus contexts

The present study focused mainly on the question of whether P3a and P3b peak parameters are modulated by perceptual references which represent independent perceptual dimensions of speech quality. To investigate this, peak amplitudes and latencies were extracted from difference epochs of specific stimulus sets and contrasted against each other. Visual inspection of the grand average difference waveforms as well as significant time points—obtained from running t-tests across the epoch time course—suggested that both P3a and P3b subcomponents were successfully triggered by the oddball task, demonstrating typical morphological and temporal characteristics (see subplot 2 in Fig. 5). Based on the same information P3a and P3b peak parameters were extracted, using separate time windows from 350 to 550 ms and from 550 to 750 ms, respectively.

Results obtained from the first step of electrophysiological data analysis supported the notion of different perceptual references for quality-impaired vs. HQ stimulus contexts due to significant modulations of peak amplitude

at electrode position Cz and peak latency at Pz. Interestingly, given C as D (see comparison 2 in Table 4), a larger P3a peak amplitude was evoked in the HQ stimulus context compared to the degraded stimulus context, the latter in this case consisting of N and P. With MOS/CMOS distances between C and N as well as C and P being larger than between C and HQ, an impaired stimulus context should have produced a larger deviation between D and S (D-S deviation) compared to a HQ stimulus context. Since the P3a constitutes a central part of the orienting response elicited by novel stimuli [61] and the occurrence of C as D was completely task-irrelevant, the opposite effect would have been expected, i.e., a larger D-S deviation causing a stronger (and faster) orienting response with larger P3a peak amplitudes (and shorter P3a peak latencies; e.g., as observed by [22, 43]). Thus, the observed differences in the neural response were attributed to the detection of change not in the quality intensity, but in the perceptual quality dimension underlying the degradation of the presented oddball stimulus.

Electrophysiological effects: perceptual references for single quality dimensions

Corresponding results emerged in the second analysis step, demonstrating that the perceptual reference affected the ERP parameters primarily within the P3a-related time window (350–550) at electrode positions Fz and Cz. For set comparisons 6 and 7 listed in Table 5, the neural response to C as D was stronger for HQ as S relative to N or P as S—again indicating a positive relationship between the size of the D-S deviation and the P3a peak amplitude, which contradicts the orienting response viewpoint. Furthermore, modulations of peak latency with regard to neural responses to T were found for set comparisons 1 and 5 (see Table 6), confirming previously reported negative effects of T-S discriminability (also termed “perceptual distinctiveness” in the literature) on P300 amplitude and latency [15, 16, 30, 48, 53]. Based on this result pattern as well as the fact that T-S discriminability directly determines the difficulty of the oddball task, an alternative account appears plausible: High T-S discriminability should allow for an easy detection task because T is clearly separable from the S sequence, hence stimulus evaluation is fast and less resources are depleted in subsequent memory-related processing as indicated by shorter P300 latencies and larger P300 amplitudes. In contrast, low T-S discriminability should make the detection task more difficult, prolonging stimulus evaluation and increasing the cognitive load of the memory operations as indicated by longer P300 latencies and smaller P300 amplitudes [31]. This explanation is further supported by post-hoc comparison 4 with N serving as T (see Table 6), which revealed a larger P300 peak

amplitude for HQ as S relative to C or P as S. Correspondingly, inspection of Fig. 6 (Subfigure 1) shows that at electrode position Cz the average neural response evoked by the presentation of T was modulated by the perceptual reference, which had been established by the previous frequent presentation of S: When N served as oddball, the P3a amplitude was highest for the HQ waveform in contrast to the C and P waveforms.

Despite the prediction of effects on P3b peak parameters, which should indicate changes in memory-related processing, all the observed modulations of P300 parameters following the occurrence of T were only significant in the P3a-related time window. Thus, it remained unclear, whether the described account would suffice as an exhaustive explanation of the observed data. Potential effects of loudness level on P3a and P3b peak parameters [46] were anticipated to be non-significant as preferred loudness levels varied across subjects and standard deviations of quality rating were very small (compare with confidence intervals plotted in Fig. 3). Again, the observed effects on P300 parameters were more likely attributable to the detection of change in the perceptual dimension of the degraded oddball stimulus, causing variation in the resource demands of memory-related processing.

Confounding factors and alternative explanations

Since the goal of the present study was to provide a first exploration of the effects of perceptual references of independent dimensions of speech quality on P3a and P3b parameters, the quality intensity of the three degraded stimuli was not controlled. This is apparent from the predicted MOS values as well as observed MOS/CMOS values, but also from the two response parameters (response time and correct rate), which were extracted from the oddball task (see again Sects. 4.1 and 4.2). Consequently, it could be argued, that the strength and speed of the elicited neural response actually depends on differences in quality intensity between the oddballs and the perceptual references detected during an early matching process, which in turn modulates the P3a and P3b parameters. Whether this alternative account is valid and how it relates to the assumed detection of changes in the underlying quality dimensions of the degraded stimuli cannot be decided based on the present experimental data, however, and therefore needs to be addressed by future studies.

It is noticeable from Table 5 that none of the comparisons with P as oddball turned out to be statistically significant. A potential explanation for this lies in the temporal domain of the degraded stimuli: In case of N and C, the degradation methods induced a continuous degradation across all samples of the audio speech file, but in case of P resulted in a discretely degraded stimulus since

only a number of distinct samples were omitted. This in turn might have led to changes both in the quality perception of the degraded stimuli and the neural responses evoked after their occurrence as oddballs: N and C were probably perceived as continuously impaired and also detected faster on the neural level, since the presented sound file was already impaired in the first sample. On the contrary, the characteristics of the neural response to P, which was presumably perceived as a discrete type of quality degradation, might have varied depending on the position of the impaired samples.

Taken together, only one significant modulation of peak latency in the time window from 550 to 750 ms could be interpreted as reflecting changes in P3b-related activity (see comparison 3 in Table 4). This is surprising, given that more T epochs than D epochs were collected as a consequence of the higher probability of T (0.2) vs. D (0.1), presumably resulting in a lower signal-to-noise ratio (SNR) for T-S vs. D-S waveforms. Since the P3a is highly sensitive to the probability of novel stimuli (oddballs)—displaying larger amplitudes and smaller latencies for more infrequent oddballs [18, 19, 52, 63, 65]—this might have led to more pronounced P3a responses after the occurrence of D. According to the theoretical framework depicted in Fig. 1, existing perceptual references should affect mostly attentional and memory-related processes located on the perception stage of the quality formation process. Since the P3a and P3b are generally associated with similar cognitive functions, i.e., P3a and P3b reflecting attentional resource allocation and memory updating operations [50, 51], they were regarded as suitable indicators of these early quality-related processes: The P3a should reflect attentional processing after the unexpected occurrence of an oddball, while the P3b should reflect memory processing after the oddball was categorized as task-relevant. Although the present study provided a first exploration of the validity of this model, the difference in probabilities precluded the addition of a second factor *oddball* to the experimental design. Hence, the assumption that P3a and P3b reflect distinct quality-related processes of attentional and memory processing could not be fully addressed in the present study. A more detailed analysis will be needed to test whether the three-stimulus oddball paradigm can reliably dissociate distinct processes in human quality perception.

Conclusions

To sum up, the analysis of the electrophysiological data provided first evidence that the three perceptual dimensions of speech quality, noisiness, coloration and discontinuity, are internally represented as different perceptual references as indicated by modulations of the P3a subcomponent. Yet,

since the degraded stimuli varied considerably in their average quality judgments (see Fig. 3), the observed effects on P300 parameters might still be ascribed to differences in degradation intensity. In a follow-up experiment, this confounding factor would have to be controlled by adjusting the degraded stimuli, such that their respective judgments on the MOS/CMOS scales are approximately equal. Degraded stimuli produced this way would then vary only in the particular perceptual dimension on which the quality degradation was initially induced.

Altogether, the present study exemplifies the use of neurophysiological methods for the assessment of speech quality, providing more direct indicators of human influencing factors on speech perception. Detailed knowledge about the internal representations and processes governing the formation of quality percepts and judgments will help to improve the evaluation, prediction and optimization of media transmission quality, ultimately increasing acceptance and long-term satisfaction on the user side [37, 38].

Acknowledgements We would like to thank the reviewers for their extensive and valuable feedback on our experimental study, which has helped to substantially improve the overall quality of the present article. Moreover, we thank Friedemann Köster and Gabriel Mittag for conducting the instrumental estimation of speech quality. We also thank Sonia Sobol for her help in designing the illustrative figures used in the article.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Antons J-N (2015) Neural correlates of quality perception for complex speech signals. Springer International Publishing, Cham. doi:10.1007/978-3-319-15521-0
2. Antons J-N, Arndt S, Schleicher R, Möller S (2014) Brain activity correlates of quality of experience. In: Möller S, Raake A (eds) Quality of experience. Springer International Publishing, Cham, pp 109–119
3. Antons J-N, Köster F, Arndt S, Möller S, Schleicher R (2013) Changes of vigilance caused by varying bit rate conditions. In: Proceedings of the 5th international workshop on quality of multimedia experience (QoMEX). IEEE, pp 148–151. doi:10.1109/QoMEX.2013.6603228
4. Antons J-N, Porbadnigk AK, Schleicher R, Blankertz B, Möller S, Curio G (2010) Subjective listening tests and neural correlates of speech degradation in case of signal correlated noise. In: Audio engineering society (AES) 129th convention. Curran Associates Inc, San Francisco

5. Antons J-N, Schleicher R, Arndt S, Möller S, Curio G (2012) Too tired for calling? A physiological measure of fatigue caused by bandwidth limitations. In: Proceedings of the 4th international workshop on quality of multimedia experience (QoMEX). IEEE, pp 63–67. doi:[10.1109/QoMEX.2012.6263840](https://doi.org/10.1109/QoMEX.2012.6263840)
6. Antons J-N, Schleicher R, Arndt S, Möller S, Porbadnigk AK, Curio G (2012) Analyzing speech quality perception using electroencephalography. *IEEE J Sel Top Signal Process* 6(6):721–731. doi:[10.1109/JSTSP.2012.2191936](https://doi.org/10.1109/JSTSP.2012.2191936)
7. Arndt S (2015) Neural correlates of quality during perception of audiovisual stimuli, 1st edn. Springer Publishing Company, Incorporated. doi:[10.1007/978-981-10-0248-9](https://doi.org/10.1007/978-981-10-0248-9)
8. Arndt S, Antons J-N, Gupta R, ur Rehman Laghari K, Schleicher R, Möller S, Falk TH (2013) The effects of text-to-speech system quality on emotional states and frontal alpha band power. In: Proceedings of the 6th international IEEE/EMBS conference on neural engineering (NER). IEEE, pp 489–492. doi:[10.1109/NER.2013.6695978](https://doi.org/10.1109/NER.2013.6695978)
9. Arndt S, Antons J-N, Schleicher R, Möller S (2016) Using electroencephalography to analyze sleepiness due to low-quality audiovisual stimuli. *Signal Process Image Commun* 42:120–129. doi:[10.1016/j.image.2016.01.011](https://doi.org/10.1016/j.image.2016.01.011)
10. Arndt S, Antons J-N, Schleicher R, Möller S, Curio G (2012) Perception of low-quality videos analyzed by means of electroencephalography. In: Proceedings of the 4th international workshop on quality of multimedia experience (QoMEX). IEEE, pp 284–289. doi:[10.1109/QoMEX.2012.6263836](https://doi.org/10.1109/QoMEX.2012.6263836)
11. Arndt S, Antons J-N, Schleicher R, Möller S, Curio G (2014) Using electroencephalography to measure perceived video quality. *IEEE J Sel Top Signal Process* 8(3):366–376. doi:[10.1109/JSTSP.2014.2313026](https://doi.org/10.1109/JSTSP.2014.2313026)
12. Arndt S, Antons J-N, Schleicher R, Möller S, Scholler S, Curio G (2011) A physiological approach to determine video quality. In: Proceedings of the IEEE international symposium on multimedia (ISM). IEEE, pp 518–523. doi:[10.1109/ISM.2011.91](https://doi.org/10.1109/ISM.2011.91)
13. Arndt S, Radun J, Antons J-N, Möller S (2014) Using eye-tracking and correlates of brain activity to predict quality scores. In: Proceedings of the 6th international workshop on quality of multimedia experience (QoMEX). IEEE, pp 281–285. doi:[10.1109/QoMEX.2014.6982332](https://doi.org/10.1109/QoMEX.2014.6982332)
14. Bakeman R (2005) Recommended effect size statistics for repeated measures designs. *Behav Res Methods* 37(3):379–384. doi:[10.3758/BF03192707](https://doi.org/10.3758/BF03192707)
15. Comerchero MD, Polich J (1998) P3a, perceptual distinctiveness, and stimulus modality. *Cognit Brain Res* 7(1):41–48. doi:[10.1016/S0926-6410\(98\)00009-3](https://doi.org/10.1016/S0926-6410(98)00009-3)
16. Comerchero MD, Polich J (1999) P3a and P3b from typical auditory and visual stimuli. *Clin Neurophysiol* 110(1):24–30. doi:[10.1016/S0168-5597\(98\)00033-1](https://doi.org/10.1016/S0168-5597(98)00033-1)
17. Duncan CC, Barry RJ, Connolly JF, Fischer C, Michie PT, Näätänen R, Petten CV (2009) Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin Neurophysiol* 120(11):1883–1908. doi:[10.1016/j.clinph.2009.07.045](https://doi.org/10.1016/j.clinph.2009.07.045)
18. Duncan-Johnson CC, Donchin E (1977) On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology* 14(5):456–467. doi:[10.1111/j.1469-8986.1977.tb01312.x](https://doi.org/10.1111/j.1469-8986.1977.tb01312.x)
19. Duncan-Johnson CC, Donchin E (1982) The P300 component of the event-related brain potential as an index of information processing. *Biol Psychol* 14(1–2):1–52. doi:[10.1016/0301-0511\(82\)90016-3](https://doi.org/10.1016/0301-0511(82)90016-3)
20. Engelke U, Darcy DP, Mulliken GH, Bosse S, Martini MG, Arndt S, Brunnström K (2016) Psychophysiology-based QoE assessment: a survey. *IEEE J Sel Top Signal Process*, pp 1–1. doi:[10.1109/JSTSP.2016.2609843](https://doi.org/10.1109/JSTSP.2016.2609843)
21. Fabiani M, Gratton G, Federmeier KD (2007) Event-related brain potentials: methods, theory, and applications. In: Cacioppo JT, Tassinary LG, Berntson G (eds) *Handbook of psychophysiology*, 3rd edn. Cambridge University Press, Cambridge, pp 85–119
22. Ford JM, Roth WT, Kopell BS (1976) Auditory evoked potentials to unpredictable shifts in pitch. *Psychophysiology* 13(1):32–39. doi:[10.1111/j.1469-8986.1976.tb03333.x](https://doi.org/10.1111/j.1469-8986.1976.tb03333.x)
23. Friedman D, Cycowicz YM, Gaeta H (2001) The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev* 25(4):355–373. doi:[10.1016/S0149-7634\(01\)00019-7](https://doi.org/10.1016/S0149-7634(01)00019-7)
24. Halgren E, Marinkovic K, Chauvel P (1998) Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalogr Clin Neurophysiol* 106(2):156–164. doi:[10.1016/S0013-4694\(97\)00119-3](https://doi.org/10.1016/S0013-4694(97)00119-3)
25. ITU-R Recommendation BS.1770-3 (2012) Algorithms to measure audio programme loudness and truepeak audio level. International Telecommunication Union, Geneva
26. ITU-T Recommendation G.722.2 (2003) Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB). International Telecommunication Union, Geneva
27. ITU-T Recommendation P.810 (1996) Modulated noise reference unit (MNRU). International Telecommunication Union, Geneva
28. ITU-T Recommendation P.863 (2011) Perceptual objective listening quality assessment. Geneva
29. Johnson R (1993) On the neural generators of the P300 component of the event-related potential. *Psychophysiology* 30(1):90–97. doi:[10.1111/j.1469-8986.1993.tb03208.x](https://doi.org/10.1111/j.1469-8986.1993.tb03208.x)
30. Katayama J, Polich J (1998) Stimulus context determines P3a and P3b. *Psychophysiology* 35(1):23–33. doi:[10.1111/1469-8986.3510023](https://doi.org/10.1111/1469-8986.3510023)
31. Kok A (2001) On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38(3):557–577
32. Kroupi E, Hanhart P, Lee JS, Rerabek M, Ebrahimi T (2014) EEG correlates during video quality perception. In: 2014 22nd European signal processing conference (EUSIPCO), pp 2135–2139
33. Lindemann L, Magnor M (2011) Assessing the quality of compressed images using EEG. In: 2011 18th IEEE international conference on image processing, pp 3109–3112. doi:[10.1109/ICIP.2011.6116324](https://doi.org/10.1109/ICIP.2011.6116324)
34. Lindemann L, Wenger S, Magnor M (2011) Evaluation of video artifact perception using eventrelated potentials. In: Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization. ACM, New York, pp 53–58. doi:[10.1145/2077451.2077461](https://doi.org/10.1145/2077451.2077461)
35. Luck SJ (2014) An introduction to the event-related potential technique, 2nd edn. The MIT Press, Cambridge
36. Möller S (2010) Quality engineering. *Qualität kommunikationstechnischer Systeme*. Springer, Berlin
37. Möller S, Raake A (eds) (2014a) Quality of experience: advanced concepts, applications and methods. Springer, Cham
38. Möller S, Raake A (2014b) Quality of experience: terminology, methods and applications. *PIK—Praxis der Informationsverarbeitung und Kommunikation*, 37(4). doi:[10.1515/pik-2014-0027](https://doi.org/10.1515/pik-2014-0027)
39. Moon S-E, Lee J-S (2015) Perceptual experience analysis for tone-mapped HDR videos based on EEG and peripheral physiological signals. *IEEE Trans Auton Ment Dev* 7(3):236–247. doi:[10.1109/TAMD.2015.2449553](https://doi.org/10.1109/TAMD.2015.2449553)
40. Mustafa M, Guthe S, Magnor M (2012) Singletrial EEG classification of artifacts in videos. *ACM Trans Appl Percept* 9(3):12:1–12:15. doi:[10.1145/2325722.2325725](https://doi.org/10.1145/2325722.2325725)
41. Mustafa M, Lindemann L, Magnor M (2012) EEG analysis of implicit human visual perception. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 513–516. doi:[10.1145/2207676.2207746](https://doi.org/10.1145/2207676.2207746)

42. Nieuwenhuis S, De Geus EJ, Aston-Jones G (2011) The anatomical and functional relationship between the P3 and autonomic components of the orienting response: P3 and orienting response. *Psychophysiology* 48(2):162–175. doi:[10.1111/j.1469-8986.2010.01057.x](https://doi.org/10.1111/j.1469-8986.2010.01057.x)
43. Novitski N, Tervaniemi M, Huotilainen M, Näätänen R (2004) Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognit Brain Res* 20(1):26–36. doi:[10.1016/j.cogbrainres.2003.12.011](https://doi.org/10.1016/j.cogbrainres.2003.12.011)
44. Nuwer MR, Comi G, Emerson R, Fuglsang-Frederiksen A, Guerit JM, Hinrichs H, Rappelsburger P (1998) IFCN standards for digital recording of clinical EEG. *International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol* 106(3):259–261
45. Olejnik S, Algina J (2003) Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychol Methods* 8(4):434–447. doi:[10.1037/1082-989X.8.4.434](https://doi.org/10.1037/1082-989X.8.4.434)
46. Papanicolaou AC, Loring DW, Raz N, Eisenberg HM (1985) Relationship Between Stimulus Intensity and the P300. *Psychophysiology* 22(3):326–329. doi:[10.1111/j.1469-8986.1985.tb01608.x](https://doi.org/10.1111/j.1469-8986.1985.tb01608.x)
47. Pizzagalli DA (2007) Electroencephalography and high-density electrophysiological source localization. In: Cacioppo JT, Tassinari LG, Berntson G (eds) *Handbook of psychophysiology*, 3rd edn. Cambridge University Press, Cambridge, pp 56–84
48. Polich J (1987) Task difficulty, probability, and interstimulus interval as determinants of P300 from auditory stimuli. *Electroencephalogr Clin Neurophysiol Evoked Potentials Sect* 68(4):311–320. doi:[10.1016/0168-5597\(87\)90052-9](https://doi.org/10.1016/0168-5597(87)90052-9)
49. Polich J (2003) Theoretical overview of P3a and P3b. In: Polich J (ed) *Detection of change*. Springer, Boston, pp 83–98. doi:[10.1007/978-1-4615-0294-4_5](https://doi.org/10.1007/978-1-4615-0294-4_5)
50. Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118(10):2128–2148. doi:[10.1016/j.clinph.2007.04.019](https://doi.org/10.1016/j.clinph.2007.04.019)
51. Polich J (2012) *Neuropsychology of P300*. Oxford University Press, Oxford. doi:[10.1093/oxfordhb/9780195374148.013.0089](https://doi.org/10.1093/oxfordhb/9780195374148.013.0089)
52. Polich J, Alexander JE, Bauer LO, Kuperman S, Morzorati S, O'Connor SJ, Begleiter H (1997) P300 topography of amplitude/latency correlations. *Brain Topogr* 9(4):275–282. doi:[10.1007/BF01464482](https://doi.org/10.1007/BF01464482)
53. Polich J, Comerchero MD (2003) P3a from visual stimuli: typicality, task, and topography. *Brain Topogr* 15(3):141–152. doi:[10.1023/A:1022637732495](https://doi.org/10.1023/A:1022637732495)
54. Polich J, Criado JR (2006) Neuropsychology and neuropharmacology of P3a and P3b. *Int J Psychophysiol* 60(2):172–185. doi:[10.1016/j.ijpsycho.2005.12.012](https://doi.org/10.1016/j.ijpsycho.2005.12.012)
55. Porbadnigk AK, Antons J, Blankertz B, Treder MS, Schleicher R, Möller S, Curio G (2010) Using ERPs for assessing the (sub) conscious perception of noise. In: *Proceeding of the annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, pp 2690–2693. doi:[10.1109/IEMBS.2010.5626549](https://doi.org/10.1109/IEMBS.2010.5626549)
56. Porbadnigk AK, Treder MS, Blankertz B, Antons J-N, Schleicher R, Möller S, Müller K-R (2013) Single-trial analysis of the neural correlates of speech quality perception. *J Neural Eng* 10(5):056003. doi:[10.1088/1741-2560/10/5/056003](https://doi.org/10.1088/1741-2560/10/5/056003)
57. Raake A, Egger S (2014) Quality and quality of experience. In: Möller S, Raake A (eds) *Quality of experience*. Springer International Publishing, Cham, pp 11–33
58. Scholler S, Bosse S, Treder MS, Blankertz B, Curio G, Müller K-R, Wiegand T (2012) Toward a direct measure of video quality perception using EEG. *IEEE Trans Image Process* 21(5):2619–2629. doi:[10.1109/TIP.2012.2187672](https://doi.org/10.1109/TIP.2012.2187672)
59. Sen D (2001) Determining the dimensions of speech quality from PCA and MDS analysis of the diagnostic acceptability measure. In: *Proceedings of MESAQUIN*, pp 1–3
60. Sen D (2004) Predicting foreground SH, SL and BNH DAM scores for multidimensional objective measure of speech quality. In: *Proceedings of IEEE international conference on acoustics, speech, and signal processing (ICASSP'04)*, vol 1, pp 493–493. doi:[10.1109/ICASSP.2004.1326030](https://doi.org/10.1109/ICASSP.2004.1326030)
61. Sokolov EN (ed) (2002) *The orienting response in information processing*. L. Erlbaum Associates, Mahwah
62. Soltani M, Knight RT (2000) Neural Origins of the P300. *Crit Rev Neurobiol* 14(3–4):26. doi:[10.1615/CritRevNeurobiol.v14.i3-4.20](https://doi.org/10.1615/CritRevNeurobiol.v14.i3-4.20)
63. Squires NK, Squires KC, Hillyard SA (1975) Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol* 38(4):387–401. doi:[10.1016/0013-4694\(75\)90263-1](https://doi.org/10.1016/0013-4694(75)90263-1)
64. Sutton S, Braren M, Zubin J, John ER (1965) Evoked-potential correlates of stimulus uncertainty. *Science* 150(3700):1187–1188. doi:[10.1126/science.150.3700.1187](https://doi.org/10.1126/science.150.3700.1187)
65. Tueting P, Sutton S, Zubin J (1970) Quantitative evoked potential correlates of the probability of events. *Psychophysiology* 7(3):385–394. doi:[10.1111/j.1469-8986.1970.tb01763.x](https://doi.org/10.1111/j.1469-8986.1970.tb01763.x)
66. Volpe U, Mucci A, Bucci P, Merlotti E, Galderisi S, Maj M (2007) The cortical generators of P3a and P3b: a LORETA study. *Brain Res Bull* 73(4–6):220–230. doi:[10.1016/j.brainresbull.2007.03.003](https://doi.org/10.1016/j.brainresbull.2007.03.003)
67. Walhovd KB, Fjell AM (2002) One-year test-retest reliability of auditory ERPs in young and old adults. *Int J Psychophysiol* 46(1):29–40. doi:[10.1016/S0167-8760\(02\)00039-9](https://doi.org/10.1016/S0167-8760(02)00039-9)
68. Wältermann M (2013) *Dimension-based quality modeling of transmitted speech*. Springer, Berlin. doi:[10.1007/978-3-642-35019-1](https://doi.org/10.1007/978-3-642-35019-1)
69. Wältermann M, Raake A, Möller S (2010) Quality dimensions of narrowband and wideband speech transmission. *Acta Acustica united with Acustica* 96(6):1090–1103. doi:[10.3813/AAA.918370](https://doi.org/10.3813/AAA.918370)
70. Wältermann M, Raake A, Möller S (2012) Direct quantification of latent speech quality dimensions. *J Audio Eng Soc* 60(4):246–254
71. Woodman GF (2010) A brief introduction to the use of event-related potentials in studies of perception and attention. *Atten Percept Psychophys* 72(8):2031–2046. doi:[10.3758/APP.72.8.2031](https://doi.org/10.3758/APP.72.8.2031)
72. Wronka E, Kaiser J, Coenen AML (2012) Neural generators of the auditory evoked potential components P3a and P3b. *Acta neurobiologiae experimentalis* 72(1):51–64