



Onset-Duration Matching of Acoustic Stimuli Revisited: Conventional Arithmetic vs. Proposed Geometric Measures of Accuracy and Precision

Björn Friedrich and Peter Heil*

Systems Physiology of Learning, Leibniz Institute for Neurobiology, Magdeburg, Germany

OPEN ACCESS

Edited by:

Laurence T. Maloney,
New York University, USA

Reviewed by:

Hinze Hogendoorn,
Utrecht University, Netherlands
Laurie Heller,
Carnegie Mellon University, USA

*Correspondence:

Peter Heil
peter.heil@lin-magdeburg.de

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: 04 August 2016

Accepted: 12 December 2016

Published: 06 January 2017

Citation:

Friedrich B and Heil P (2017)
Onset-Duration Matching of Acoustic
Stimuli Revisited: Conventional
Arithmetic vs. Proposed Geometric
Measures of Accuracy and Precision.
Front. Psychol. 7:2013.
doi: 10.3389/fpsyg.2016.02013

Onsets of acoustic stimuli are salient transients and are relevant in humans for the perception of music and speech. Previous studies of onset-duration discrimination and matching focused on whether onsets are perceived categorically. In this study, we address two issues. First, we revisit onset-duration matching and measure, for 79 conditions, how accurately and precisely human listeners can adjust the onset duration of a comparison stimulus to subjectively match that of a standard stimulus. Second, we explore measures for quantifying performance in this and other matching tasks. The conventional measures of accuracy and precision are defined by arithmetic descriptive statistics and the Euclidean distance function on the real numbers. We propose novel measures based on geometric descriptive statistics and the log-ratio distance function, the Euclidean distance function on the positive-real numbers. Only these properly account for the fact that the magnitude of onset durations, like the magnitudes of most physical quantities, can attain only positive real values. The conventional (arithmetic) measures possess a convexity bias that yields errors that grow with the width of the distribution of matches. This convexity bias leads to misrepresentations of the constant error and could even imply the existence of perceptual illusions where none exist. This is not so for the proposed (geometric) measures. We collected up to 68 matches from a given listener for each condition (about 34,000 matches in total) and examined inter-listener variability and the effects of onset duration, plateau duration, sound level, carrier, and restriction of the range of adjustable comparison stimuli on measures of accuracy and precision. Results obtained with the conventional measures generally agree with those reported in the literature. The variance across listeners is highly heterogeneous for the conventional measures but is homogeneous for the proposed measures. Furthermore, the proposed measures show that listeners tend to under- rather than to overestimate the onset duration of the comparison stimuli. They further reveal effects of the stimulus carrier on accuracy and precision which are missed by the conventional measures. Our results have broad implications for psychophysical studies that use arithmetic measures to quantify performance when geometric measures should instead be used.

Keywords: psychophysics, method of adjustment, positive-real numbers, log-ratio metric, convexity bias, illusions, constant error, Weber fraction

1. INTRODUCTION

Onsets are salient features of the temporal envelopes of acoustic stimuli, with perceptual and behavioral relevance in a variety of species from insects to humans. For example, the onset duration, also called rise time or attack time, of sound pulses modeling songs of female grasshoppers determines whether a male turns toward the sound source (von Helversen, 1993). Similarly, phonotactic responses of female tree frogs to synthetic male mating calls depend on the onset duration of the calls (Gerhardt and Schul, 1999). In humans, characteristics of the onset influence the timbre of music and speech sounds (e.g., McAdams et al., 1995; Lakatos, 2000; Halpern et al., 2004). For instance, the auditory perception of a trumpet can be transformed into that of a violin, and vice versa, by adequately manipulating the properties of the sound signal during the initial 50 ms or so (Grey and Gordon, 1978). That the onset is crucial for the identification of an instrument is also supported by experiments showing that when instrumental sounds with asymmetrical onsets and offsets are played backward—a manipulation that does not alter the long-term spectrum—human listeners often fail to recognize the instrument (Paquette and Peretz, 1997). In speech, studies of general grouping principles of speech components imply that the onset of a sound's envelope is more important than what follows (Darwin, 1981).

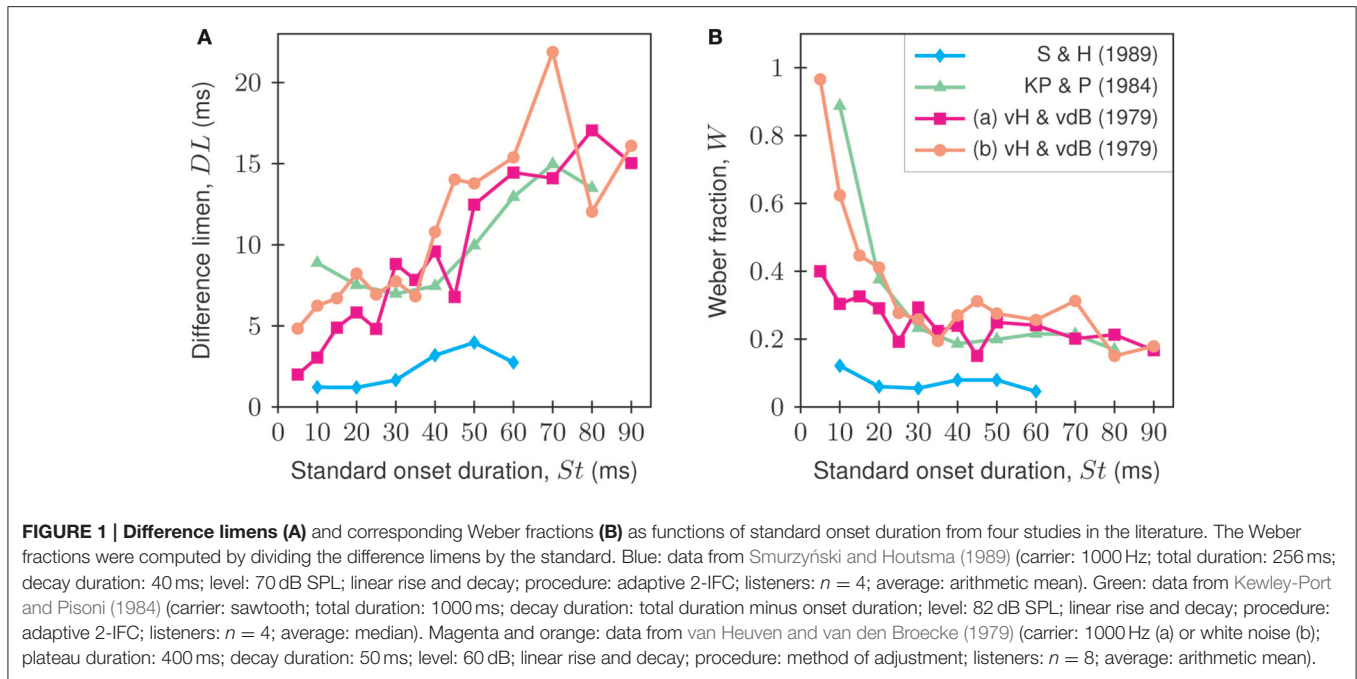
Cutting and Rosner (1974, 1976) and Cutting (1982) suggested that non-speech sounds varying along an onset-duration continuum are perceived categorically and that onset duration can therefore cue categorical perception, such as that of music and speech sounds. Rosen and Howell (1981) showed, however, that the peak in the discrimination performance observed by Cutting and Rosner (1974, 1976) was due entirely to the stimuli not having the intended onset durations. For stimuli having the intended onset durations, discrimination performance decreased monotonically with increasing onset duration (Rosen and Howell, 1981, 1983). Other investigators also questioned the categorical perception of the onset-duration continuum and suggested that the ability to discriminate or match onset durations rather follows Weber's law (Tenney, 1962; Pollack, 1963; van Heuven and van den Broecke, 1979; Hary and Massaro, 1982; van den Broecke and van Heuven, 1983; Kewley-Port and Pisoni, 1984; Smurzyński, 1985; Smurzyński and Houtsma, 1989), which states that the Weber fraction is constant.

Reported Weber fractions vary widely among studies—between 6 and 8% (Smurzyński and Houtsma, 1989) and 100% or even more (Pollack, 1963). However, Weber fractions are approximately constant over only a limited range of onset durations. Below this range, they consistently increase with decreasing onset duration (Tenney, 1962; Pollack, 1963; van Heuven and van den Broecke, 1979; van den Broecke and van Heuven, 1983; Kewley-Port and Pisoni, 1984). Moreover, Weber fractions may decrease with training (Smurzyński and Houtsma, 1989), and they depend on stimulus level. Kalmylova and Shakhshaev (1988) showed, for a single onset duration of 15 ms, a decrease in the Weber fraction with increasing sound level. **Figure 1** shows difference limens and corresponding Weber fractions reported in four studies. The small Weber fractions

from Smurzyński and Houtsma (1989), who used tone carriers and a sound level of 70 dB SPL, are in line with the finding of Kalmylova and Shakhshaev (1988) that Weber fractions are smallest (onset-duration discrimination is best) at high sound levels. Weber fractions reported by Kewley-Port and Pisoni (1984), who used broadband sawtooth carriers, are larger and increase more rapidly with decreasing onset duration below about 30 ms than those observed by Smurzyński and Houtsma (1989), even though the former used an even higher sound level (82 dB SPL). It therefore seems that the type of carrier might also affect the Weber fraction. An influence of the carrier can also be inferred from the study of van Heuven and van den Broecke (1979), who obtained somewhat smaller Weber fractions for stimuli with tone than with broadband noise carriers of the same sound level (60 dB), though the data are noisy.

Van Heuven and van den Broecke (1979) and van den Broecke and van Heuven (1983) used the method of adjustment, a fundamental and classical psychophysical procedure devised by Fechner (1860). In this method, listeners adjust a variable comparison stimulus (Co) until it subjectively equals a fixed standard stimulus (St) serving as a reference. Ideally, this is repeated many times, so that for each St there is a distribution of comparison stimuli that the listener accepted as a match. The center of this distribution is the point of subjective equality (PSE), and the signed distance between the PSE and the St is the constant error (CE), also referred to as the systematic error. The more accurate the adjustments are, the smaller the magnitude of the CE. The sign indicates the direction of the CE. A measure of the width of the distribution is used to define the difference limen (DL), conceived of as the minimum amount the stimulus must change to produce a just-noticeable difference (JND) in the sensation. The more precise the adjustments are, the smaller the DL. Relative precision is often quantified by the Weber fraction (W), commonly defined as the ratio of the DL to the St, $W = DL/St$. The listeners in the study of van Heuven and van den Broecke (1979) made only two matches per condition, and difference limens and Weber fractions were calculated from matches pooled across all listeners. Moreover, constant errors were not reported, although, as we will show in the Results section, constant errors and Weber fractions may interact.

We therefore believe that a thorough reexamination of onset-duration matching with respect to stimulus factors affecting the accuracy and precision of adjustments, as well as with respect to inter-listener variability, is in order. To reexamine onset-duration matching, we performed three experiments with up to ten listeners each. In Experiment 1, we explored the dependence of accuracy and precision on onset duration (varying from 0.5 to 64 ms), sound level (varying over a 24-dB range of low sensation levels), and carrier (tone and noise). In Experiment 2, we examined whether accuracy and precision depend on plateau duration, which covaried with onset duration in Experiment 1. In Experiment 3, we examined how restricting the range of comparison stimuli affects accuracy and precision. In all experiments, each listener performed many matches for each condition, allowing us to characterize the distributions of matches, to derive reliable estimates from each listener and condition, and to explore inter-listener variability.



The second major focus of this study concerns the way in which accuracy and precision should be quantified. We believe that the statistical measures used to quantify accuracy and precision in tasks using the method of adjustment deserve reexamination. Matches to a given standard are typically conceived of as realizations of a random variable, X_{St} , with distributions over the Euclidean standard vector space \mathbb{R} . In this space, vector addition and scalar multiplication are defined as usual addition and multiplication, respectively, and the Euclidean distance is given by the function

$$\text{dist}_{\mathbb{R}}(x, y) = |x - y| . \tag{1}$$

For convenience, we hereafter refer to this additive structure simply as \mathbb{R} . In line with the additive structure of \mathbb{R} , the textbook definitions (see e.g., Guilford, 1954; Gescheider, 1997; Ramajanickam, 2002; Mapp and Ono, 2006) of the four psychophysical measures, PSE, CE, DL, and W, are computed from X_{St} using arithmetic descriptive statistics such as the arithmetic mean $\text{AM}(X_{St})$, arithmetic variance $\text{AV}(X_{St})$, and arithmetic standard deviation $\text{AS}(X_{St})$. These conventional (arithmetic) psychophysical measures are summarized in **Table 1**.

However, the magnitudes of most fundamental (e.g., mass, length, duration) and many derived (e.g., area, volume, density, pressure, frequency) physical quantities can attain only positive real values (BIPM, 2006; Gupta, 2010). Even though the set of positive-real numbers, $\mathbb{R}_{>0}$, is a subset of the set of real numbers, \mathbb{R} , it is not a subspace in terms of the vector space structure of \mathbb{R} . Instead, $\mathbb{R}_{>0}$ has its own Euclidean vector space structure in which multiplication and exponentiation naturally take on the role of vector addition and scalar multiplication. Again, for convenience, we hereafter refer to this multiplicative structure

simply as $\mathbb{R}_{>0}$. The Euclidean distance, the metric induced by the standard-scalar-product norm, is given by the log-ratio distance function

$$\text{dist}_{\mathbb{R}_{>0}}(x, y) = \left| \ln\left(\frac{x}{y}\right) \right| = |\ln(x) - \ln(y)| . \tag{2}$$

This metric measures relative rather than absolute differences between stimuli and has a number of other interesting properties (e.g., Graff, 2014). Consistent with the structure of $\mathbb{R}_{>0}$ are the geometric mean $\text{GM}(X_{St})$, geometric variance $\text{GV}(X_{St})$, and geometric standard deviation $\text{GS}(X_{St})$ of the random variable X_{St} over $\mathbb{R}_{>0}$. Formulae to calculate the arithmetic and the geometric statistical measures are summarized in Table 1 of the Supplementary Material. The most important operations and properties of the two Euclidean vector spaces \mathbb{R} and $\mathbb{R}_{>0}$ are compared in Table 2 of the Supplementary Material.

It is important to realize that arithmetic descriptive statistics possess a convexity bias when computed from random variables over the multiplicative Euclidean vector space of positive-real numbers, and that this convexity bias is present in the conventional psychophysical measures. This can be demonstrated in a straightforward manner with the following Gedanken experiment. Let X_{St} follow a log-normal distribution, also called a multiplicative normal distribution (Limpert et al., 2001; Limpert and Stahel, 2011)—the canonical probability distribution with support on the Euclidean vector space $\mathbb{R}_{>0}$. We hereafter use the terms *multiplicative normal* and *additive normal* for the log-normal and the normal distributions, respectively. Let us further assume the special case that the subject is perfectly accurate, such that the center of the multiplicative normal distribution is exactly equal to the standard, i.e., $\text{GM}(X_{St}) = St$. The logarithm of X_{St} then follows an additive normal distribution

(\mathcal{N}) with mean $\mu = \ln(St)$ and variance σ^2 , such that $\ln(X_{St}) \sim \mathcal{N}(\mu, \sigma^2)$. Using known relationships between statistical descriptive measures and the parameters of a multiplicative normal distribution (see Table 3 of the Supplementary Material), one can reformulate the famous inequality $AM(X_{St}) \geq GM(X_{St})$, a special case of Jensen's inequality $\ln(AM(X_{St})) \geq AM(\ln(X_{St}))$, to obtain an equality with a known bias factor, namely the square root of the geometric variance:

$$AM(X_{St}) = GM(X_{St}) \cdot \sqrt{GV(X_{St})}. \quad (3)$$

The name convexity bias originates from the fact that the logarithm is a negatively convex function. The impact of the convexity bias on the conventionally defined PSE and CE is illustrated in **Figure 2**. The PSE, being defined as the arithmetic mean, differs from the St (being equal to the geometric mean) by a factor equal to $\sqrt{GV(X_{St})}$ (**Figure 2A**). Because this factor is always greater than 1, the use of the conventional measures suggests an absolute CE that is positive and grows linearly with St , the more so the wider the distribution (**Figure 2B**):

$$\text{abs. } CE_A = PSE_A - St = \left(\sqrt{GV(X_{St})} - 1\right) \cdot St > 0. \quad (4)$$

The relative CE, defined as $\text{rel. } CE_A = \frac{\text{abs. } CE_A}{St}$, is therefore also positive. It is equal to the slopes of the lines in **Figure 2B** and independent of St . The use of CE_A as a measure of accuracy therefore implies the existence of positive deviations from St even though the subject was perfectly accurate; the positive values of CE_A are of purely mathematical origin. This is important to realize in, for example, studies of perceptual illusions. When positive values of CE_A are obtained, it is not justified to infer the existence of an illusion or bias, because the positive values may be a result of the convexity bias. The measure of accuracy that we propose is the geometric constant error CE_G , which we derive in the Results section as the geometric counterpart of $\text{rel. } CE_A$. In the special case assumed in the Gedanken experiment presented above, it yields a value of $CE_G = \ln\left(\frac{GM(X_{St})}{St}\right) = 0$ (not shown in **Figures 2A,B**), which correctly indicates that the subject was perfectly accurate.

In the more general case, a subject may not be perfectly accurate. The center of the multiplicative normal distribution in our Gedanken experiment would then be $GM(X_{St}) = St \cdot e^{\mu_0}$, where e^{μ_0} represents the positive error factor by which the center deviates from St . Therefore, $\ln(X_{St}) \sim \mathcal{N}(\mu + \mu_0, \sigma^2)$ with $\mu = \ln(St)$. Our proposed measure of accuracy yields a value of $CE_G = \ln\left(\frac{GM(X_{St})}{St}\right) = \mu_0$. As shown in **Figure 2C**, $\text{rel. } CE_A$ and CE_G can both be positive (quadrant I) or both be negative (quadrant III), but it is also possible that $\text{rel. } CE_A$ is positive while CE_G is negative (quadrant II). The $\text{rel. } CE_A$ does not fulfill the purpose of a measure of accuracy when data follow a multiplicative normal distribution, because it does not always reflect the true sign and never reflects the true magnitude of the deviation from St .

The second major goal of our study was, therefore, to derive novel definitions of the psychophysical measures using geometric descriptive statistics and the log-ratio distance function $\text{dist}_{\mathbb{R}_{>0}}$

TABLE 1 | Conventional and proposed psychophysical measures, computed from the matches to a given standard (St), realized as a random variable X_{St} .

Psychophysical measure	Conventional (arithmetic)	Proposed (geometric)
Point of subjective equality	$PSE_A = AM(X_{St})$	$PSE_G = GM(X_{St})$
Absolute constant error	$\text{abs. } CE_A = PSE_A - St$	—
Relative constant error [†]	$CE_A = \frac{PSE_A - St}{St}$	$CE_G = \ln\left(\frac{PSE_G}{St}\right)$
Difference limen	$DL_A = AS(X_{St})$	$DL_G = GS(X_{St})$
Weber fraction	$W_A = \frac{DL_A}{St}$	$W_G = \ln(DL_G)$

The conventional measures are based on arithmetic descriptive statistics (subscript "A") and the distance function $\text{dist}_{\mathbb{R}}$ of the Euclidean vector space \mathbb{R} . Our proposed measures are based on geometric descriptive statistics (subscript "G") and the log-ratio distance function $\text{dist}_{\mathbb{R}_{>0}}$ of the Euclidean vector space $\mathbb{R}_{>0}$. The measures are derived in the Results section.

[†]Note that we write CE_A instead of $\text{rel. } CE_A$.

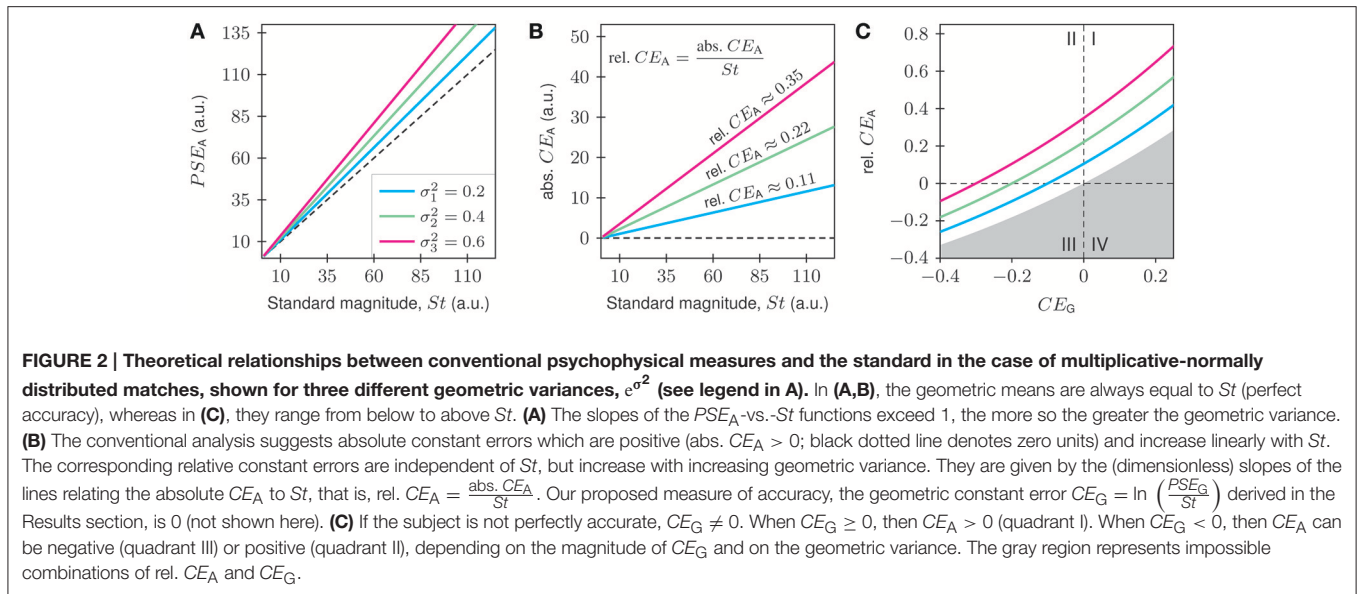
that are consistent with positive-real physical quantities and that are not affected by the convexity bias. Other potential sources of bias, such as deviations of the distributions of matches from the multiplicative normal distribution, stimulus order, memory or attentional effects, or true perceptual illusions, may still remain and be reflected by the novel measure of accuracy. The proposed, geometric measures are summarized in **Table 1** along with the conventional, arithmetic measures.

The third major goal of our study was to compare the outcomes of the analyses of our data using the conventional and the proposed novel definitions. We will show that the proposed definitions yield more homogeneous results and reveal effects of the carrier on accuracy and precision that are missed by the conventional definitions. We also find for our data that the conventional and proposed measures of accuracy have opposite signs.

2. METHODS

2.1. Listeners

Twenty-two listeners participated in the study, with ten in Experiment 1 (six male and four female, ages 23–28 years), four in Experiment 2 (two male and two female, ages 25–29 years), and eight in Experiment 3 (four male and four female, ages 21–29 years). Listeners were informed about the purpose of the study, gave their written consent to participate, and were remunerated on an hourly basis. Twelve listeners had participated in other experiments in our lab during which detection and discrimination thresholds were measured using forced-choice procedures, but none had participated in experiments using the method of adjustment. For each listener, audiograms were measured prior to the experiments by means of standard audiometry (Otometrics Madsen Itera II) and found to be normal for both ears. The study was approved by the ethics committee of the Otto von Guericke University, Magdeburg.



2.2. Stimuli

Each acoustic stimulus consisted of an onset during which the amplitude of the stimulus rose linearly from zero to the maximum, a plateau during which the amplitude remained constant, and an offset during which the amplitude decayed linearly from the maximum to zero. Depending on the experiment, stimuli could differ in four attributes: onset duration, total duration, carrier, and sound level. They were generated in MATLAB R2010a (The Mathworks, Natick, MA, USA) and stored in the Waveform Audio File Format (WAVE) for replay by custom-made stimulus-presentation software. The system was calibrated using defined sound sources (Microtech Gefell MG 4000 and Brüel and Kjaer pistonphone 4228) and an artificial ear (Brüel and Kjaer 4153) equipped with a condenser microphone (Brüel and Kjaer 4133) connected to a conditioning amplifier (Brüel and Kjaer Nexus) and a multimeter (Agilent 34401A). During presentation, stimulus levels were controlled by a programmable attenuator (PA5, Tucker-Davis Technologies) linked downstream to a 24-bit sound card (Juli@, ESI Audiotechnik GmbH) of a desktop computer.

2.2.1. Experiment 1

The onset durations (rise times) of the standard stimuli were between 0.5 and 64 ms, and neighboring values differed by a factor of 2. The onset durations of the comparison stimuli were between 0.25 and 128 ms long, and neighboring values differed by a factor of $2^{1/6} \approx 1.1225$. The offset duration of each stimulus was 50 ms. The total duration of each stimulus was 400 ms such that the plateau duration covaried with the onset duration. Two different carriers were used, a 3125-Hz tone and frozen white Gaussian noise. The stimuli were presented at sensation levels of 10, 22, and 34 dB SL.

2.2.2. Experiment 2

Total stimulus durations were 200, 400, or 800 ms, including an offset duration of 50 ms. Onset durations of the standard stimuli

were 2 ms and 64 ms. As in Experiment 1, the onset durations of the comparison stimuli were between 0.25 and 128 ms, and neighboring values differed by a factor of $2^{1/6} \approx 1.1225$. The carriers were a 3125-Hz tone and frozen white Gaussian noise. The stimuli were presented at sensation levels of 10 and 34 dB SL.

2.2.3. Experiment 3

The total duration of all stimuli was 480 ms, including an offset duration of 48 ms. The onset duration of all standard stimuli was 10.24 ms. The possible onset durations of the comparison stimuli depended on the degree of range restriction. The smallest degree of restriction (degree 1) corresponded to the broadest range from 0.64 to 163.84 ms. The largest degree of restriction (degree 7) yielded the narrowest range from 6.64 to 15.79 ms. For all ranges (see Table 2), neighboring values differed by a factor of $2^{1/16} \approx 1.0443$. The carrier was a 3125-Hz tone. The stimuli were presented at a sensation level of 30 dB SL.

2.3. Procedures

2.3.1. General Procedures

Listeners sat on a comfortable chair within an illuminated and ventilated doubled-walled soundproof chamber (IAC Acoustics). Their task was to adjust a comparison stimulus to match a standard stimulus. In each trial of an experimental condition, a pair of stimuli consisting of the standard stimulus and the comparison stimulus, separated by a silent interval of 300 ms, was presented diotically using circumaural headphones (Sennheiser HDA 200). In parallel, a computer screen displayed a window with four buttons: one for increasing and one for decreasing the onset duration of the comparison stimulus, one for reporting a match, and one for starting, pausing, and continuing the experiment. Depending on individual preference, listeners could use a computer mouse or a keyboard to activate the buttons.

In each trial in each experiment, standard and comparison stimuli had the same carrier, the same total duration, and the same sound level, but differed in onset duration. In Experiments

1 and 2, the onset duration of the standard stimulus was selected at random from the list without replacement. The onset duration of the comparison stimulus in the first trial was selected at random from a list containing the products of the standard onset durations and the factors between $2^{-12/6}$ and $2^{-6/6}$ and between $2^{+6/6}$ and $2^{+12/6}$. In Experiment 3, the onset duration of the first comparison stimulus was selected at random from ranges below and above the standard, as specified in **Table 2**. In each experiment, the listener could use the up- and down-arrow buttons, either once or multiple times in quick succession, to increase or decrease the onset duration of the comparison stimulus to be presented in the next trial. Each button press altered the onset duration by a factor of $2^{\pm 1/6}$ (in Experiments 1 and 2) or $2^{\pm 1/16}$ (in Experiment 3). The standard was not changed. Listeners could repeat this adjustment process as often as they liked. When they perceived the standard and comparison stimuli to be the same, they indicated it with a button press. No feedback was provided. Next, a different onset duration for the standard stimulus (Experiments 1 and 2) or degree of range restriction (Experiment 3; see **Table 2**) was selected and the procedure was repeated. The random selection of the initial onset duration of the comparison stimulus from the ranges specified above made it unlikely that listeners adopted simple procedural strategies to solve the task.

To complete a given experiment, every listener had to come into the lab for multiple sessions over several months. A session lasted 2–2.5 h, including one or two breaks, as desired by the listener. Before the first training session started, the listeners' audiograms were measured. In Experiments 1 and 2, each listener underwent two guided training sessions with feedback to become familiar with the task and the stimuli. Detection thresholds for tone and noise stimuli were also measured, using a modified staircase procedure. The threshold measurements were repeated at the beginning of every fifth experimental session, because thresholds may vary from day to day (Hempstock et al., 1966; Heil et al., 2006). The thresholds were used to determine the attenuation required to achieve the desired sensation levels. Level adjustments were made when necessary. In Experiment 3, each listener underwent one guided training session with feedback prior to the first experimental session. Each subsequent session began with a training block to remind listeners of the task. The experimental blocks were preceded by detection-threshold measurements for tone stimuli, using a 3-alternative-forced-choice procedure. The thresholds were used to determine the attenuation required to achieve the desired sensation level of 30 dB SL.

2.3.2. Structure of Sessions

2.3.2.1. Experiment 1

A session in Experiment 1 was constructed as follows. The sensation level of the initial standard stimulus was 34 dB SL, to start the session with the most clearly audible stimuli. The carrier (tone or noise) of the initial standard stimulus was selected at random to minimize order effects. The onset duration was also selected at random from the list of eight values (0.5, 1, 2, 4, 8, 16, 32, and 64 ms). The onset duration for the next standard stimulus was then drawn from the remaining seven values, and so on, until

TABLE 2 | Ranges for the comparison stimuli used in Experiment 3.

Degree	Total range (ms)	Initial range < St (ms)	Initial range > St (ms)
1	0.64–163.84	2.56–5.12	20.48–40.96
2	0.95–110.94	3.04–5.58	18.78–34.44
3	1.40–75.12	3.78–6.36	16.49–27.74
4	2.06–50.87	4.50–6.93	15.12–23.32
5	3.04–34.44	5.58–7.56	13.87–18.78
6	4.50–23.32	6.64–8.25	12.72–15.79
7	6.64–15.79	8.25–9.39	11.17–12.72

In Experiment 3, the degree of range restriction determined the total range of adjustable onset durations and the initial ranges below and above the onset duration of the St (10.24 ms) from which the onset duration of the Co was randomly selected on the first trial.

all eight standard onset durations had been drawn once, without changing the carrier or the sound level. Then, the next lower sound level (22 dB SL) was selected and the procedure repeated until all standard onset durations had been drawn once. Finally, the lowest sound level (10 dB SL) was selected and the procedure repeated until all standard onset durations had been drawn once. The entire procedure was then repeated for the other carrier, using all combinations of onset durations and sound levels. The entire procedure was repeated once more for each carrier, starting with the first. In this way, a listener made 96 adjustments in a single experimental session, with two adjustments for each of the 48 conditions (8 standard onset durations \times 3 sound levels \times 2 carriers). Each of the 10 listeners completed between 18 and 31 such sessions and therefore made between 36 and 62 adjustments for each of the 48 conditions (between 1728 and 2976 adjustments total).

2.3.2.2. Experiment 2

A session in Experiment 2 was constructed as follows. The sensation level of the initial standard stimulus was 34 dB SL, to start the session with the most clearly audible stimuli. The carrier (tone or noise) of the initial standard stimulus was selected at random to minimize order effects. The onset duration (2 or 16 ms) and total duration (200, 400, or 800 ms) were also selected at random. The remaining onset duration was used for the next standard stimulus, while the total duration was unchanged. After both onset durations had been drawn once, the next total duration was selected, and so on, until all six combinations of onset and total durations had been presented four times. This whole procedure was then repeated for the sound level of 10 dB SL. The listener was encouraged to take a break of about 10 min, after which the whole procedure was repeated for the other carrier. In this way, a listener made 96 adjustments in a single experimental session, with four adjustments for each of the 24 conditions (2 standard onset durations \times 3 total durations \times 2 carriers \times 2 sound levels). Each of the four listeners completed between 11 and 17 such sessions and therefore made between 44 and 68 adjustments for each condition (between 1056 and 1632 adjustments total).

2.3.2.3. Experiment 3

A session in Experiment 3 began with a training block. Each experimental block was preceded by the detection-threshold

measurements. An experimental block consisted of three or four repetitions of the seven randomly permuted degrees of range restriction, such that one experimental block consisted of 21 or 28 trials. The listeners performed as many experimental blocks as possible (usually four) within about 2 h of measurement time. Each of the eight listeners made between 74 and 110 adjustments for each of the seven degrees of range restriction (between 518 and 770 adjustments total).

2.4. Data Analysis

All data analyses were performed using MATLAB R2010a (The Mathworks, Natick, MA, USA) with the Curve Fitting Toolbox, the Optimization Toolbox, the Parallel Computing Toolbox, and the Statistics Toolbox. In all statistical tests the α -criterion was 0.05.

3. RESULTS

3.1. Onset-Duration Matches Are Consistent with a Multiplicative Model

According to our theoretical considerations, onset durations should be treated as elements of the multiplicative Euclidean vector space $\mathbb{R}_{>0}$ rather than of the additive Euclidean vector space \mathbb{R} . To investigate empirically whether the matches of onset durations are consistent with an additive or with a multiplicative statistical model, we used the maximum-likelihood method to fit additive normal and multiplicative normal distributions to the distributions of matches obtained in Experiments 1 and 2, separately for each condition and listener (576 distributions in total, each based on between 36 and 68 matches). Even if other distributions might fit a particular dataset better than the additive normal and multiplicative normal distributions do, these two represent the simplest distributions describing data that behave according to an additive or a multiplicative Gaussian error model (Limpert et al., 2001; Limpert and Stahel, 2011). In fact, they are the maximum-entropy distributions when only the mean and the variance are known (Park and Bera, 2009). Fitting these two distributions therefore suffices to distinguish between additive and multiplicative models as long as no other constraints are imposed. The results of the fits are summarized in **Figure 3**.

Figure 3A shows, for one example listener in Experiment 1, the matches (blue dots; $n_{\text{mat}} = 62$ per St) for each of the eight standard onset durations (gray vertical lines) of tone stimuli presented at 34 dB SL. Superimposed are the cumulative distribution functions (CDFs) of the fitted multiplicative normal distributions (black lines). They fit the data well.

Figure 3B is a scatterplot of the negative log-likelihoods from the fits to the multiplicative normal distribution, ℓ_{mult} , and to the additive normal distribution, ℓ_{add} , for all conditions and listeners in Experiments 1 and 2. The data from the example listener and conditions in **Figure 3A** are indicated separately. The smaller ℓ , the better the fit. In 471 of 576 cases (82%), the multiplicative normal distribution fits the data better than the additive normal distribution. This is illustrated more clearly in the inset of **Figure 3B**, which shows the cumulative distribution of the differences between the negative log-likelihoods, $\Delta\ell = \ell_{\text{mult}} - \ell_{\text{add}}$. Negative values of $\Delta\ell$ support the multiplicative normal

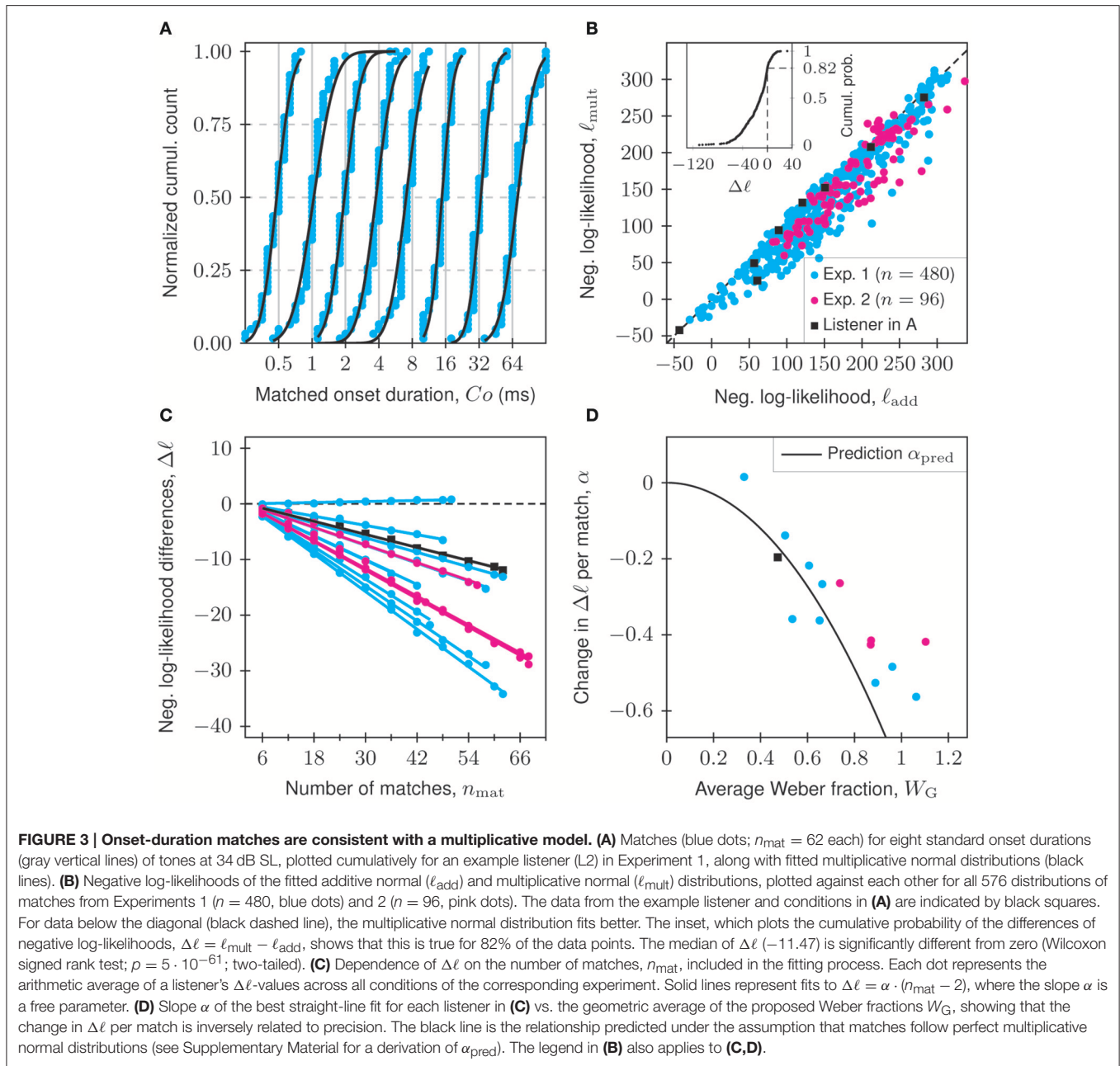
distribution, whereas positive values support the additive normal distribution. Dashed lines indicate the fraction of negative differences (0.82). The median of $\Delta\ell$ (-11.47) is significantly different from zero (Wilcoxon signed rank test; $p = 5 \cdot 10^{-61}$; two-tailed).

Because previous studies have obtained only few matches per listener (e.g., two matches in van Heuven and van den Broecke, 1979; van den Broecke and van Heuven, 1983), we also tested how the number of matches included in the fits influences the support for each model. From the distribution of matches for each condition and listener, we randomly drew a number of matches ($n_{\text{mat}} \in \{6, 12, 18, 24, \dots\}$ or all), fitted them with additive normal and multiplicative normal distributions using the maximum likelihood method, and computed $\Delta\ell$. To reduce the noise, we arithmetically averaged the $\Delta\ell$ across all conditions for a given n_{mat} and listener. **Figure 3C** plots these averages of $\Delta\ell$ (dots) as a function of n_{mat} , separately for each of the 14 listeners in Experiments 1 and 2. For all listeners but one, the multiplicative normal distribution is increasingly favored as additional matches are included. This shows that a sufficient number of matches per listener and condition is required for a clear empirical differentiation between the additive and the multiplicative models. Superimposed in **Figure 3C** (solid lines) are least-squares fits of the data from each listener to the straight-line equation $\Delta\ell = \alpha \cdot (n_{\text{mat}} - 2)$ with $n_{\text{mat}} \geq 2$. The offset was fixed to take into account that $\Delta\ell = 0$ for $n_{\text{mat}} = 2$. The slope α , which represents the change in $\Delta\ell$ per match included in the fit, was a free parameter. For each listener, the slope is plotted in **Figure 3D** against the geometric average across all conditions of the geometric Weber fraction, W_G (see derivation in next section). The slopes are shallow for listeners with small average W_G (narrow distributions; high precision) and steep for listeners with large average W_G (broad distributions; low precision). The black line shows the relationship predicted under the assumption that all matches follow perfect multiplicative normal distributions. The data points fall near, though mostly above or to the right of, the prediction. Simulations (not shown) revealed that this is partly explained by averaging $\Delta\ell$ and W_G from different distributions. In addition, the experimental data do not follow perfect multiplicative normal distributions.

In summary, when a sufficient number of matches is included, clear empirical support for the multiplicative model is obtained. This corroborates our theoretical considerations and the conclusion that onset-duration adjustment data should be described using geometric statistics and novel psychophysical measures, rather than arithmetic statistics and conventional psychophysical measures used in previous studies (e.g., van Heuven and van den Broecke, 1979; van den Broecke and van Heuven, 1983; Smurzyński and Houtsma, 1989). The novel measures are derived in the next section.

3.2. Derivation of Psychophysical Performance Measures for Positive-Real Quantities

Here, we derive probabilistic definitions of the psychophysical performance measures for the method of adjustment when the



physical stimuli used can have only positive real magnitudes. We base our derivation on the (one-dimensional) Euclidean vector space $\mathbb{R}_{>0}$ in which multiplication and exponentiation take on the role of vector addition and scalar multiplication and distances between any two elements $x, y \in \mathbb{R}_{>0}$ are measured with the log-ratio distance function $\text{dist}_{\mathbb{R}_{>0}}(x, y) = |\ln(x) - \ln(y)| = \left| \ln\left(\frac{x}{y}\right) \right|$. To obtain the operations and properties of $\mathbb{R}_{>0}$, one can exploit the fact that \mathbb{R} and $\mathbb{R}_{>0}$ are isometrically isomorphic, meaning that there are bijective functions from one space to the other that preserve the algebraic and metric structure. The most natural choice of bijections is the exponential function $\exp: \mathbb{R} \rightarrow \mathbb{R}_{>0}, x \mapsto e^x$ and the logarithmic function $\ln: \mathbb{R}_{>0} \rightarrow \mathbb{R}, x \mapsto$

$\log_e(x)$, where $e \approx 2.71828$ is the Euler-Napier number. The Euclidean vector spaces \mathbb{R} and $\mathbb{R}_{>0}$ are compared in Table 2 of the Supplementary Material.

For the method of adjustment, we are interested in calculating, for a given standard magnitude St , the constant and variable errors of a distribution of matches with positive-real magnitudes. To simplify matters, we assume that X_{St} approximately follows the canonical Gaussian distribution on $\mathbb{R}_{>0}$, namely the multiplicative normal distribution, for which the relationships between parameters and statistical measures are well known (see Table 3 of the Supplementary Material). Let μ_0 and σ^2 be the parameters representing a potential constant error and the

variable error, respectively, and let $\mu = \ln(St)$. Note that when X_{St} follows a multiplicative normal distribution, the logarithm of X_{St} follows an additive normal distribution with the same parameters, i.e., $\ln(X_{St}) \sim \mathcal{N}(\mu + \mu_0, \sigma^2)$. Conceptually, the point of subjective equality should be the center of X_{St} , such that

$$PSE_G = e^{\mu + \mu_0} = St \cdot e^{\mu_0} = GM(X_{St}) . \tag{5}$$

We propose to define the constant error as the directed log-ratio distance between the PSE and the St:

$$CE_G = \ln\left(\frac{PSE_G}{St}\right) = \ln(PSE_G) - \ln(St) = \mu_0 . \tag{6}$$

This directed distance differs from $\text{dist}_{\mathbb{R}_{>0}}(PSE_G, St)$ in that the sign of the error is retained. It is a dimensionless measure with the logarithmic auxiliary unit neper (Np), which can be converted to the more familiar decibel (dB), where $1 \text{ Np} = (20/\ln(10)) \text{ dB} \approx 8.686 \text{ dB}$. By analogy with the Euclidean space \mathbb{R} , where the magnitude of the comparison stimulus that is just-noticeably different from the St is typically chosen to be $Co = St \pm AS(X_{St})$, let us define

$$Co = St * GS(X_{St}) , \tag{7}$$

where the symbol $*$ denotes “times or divided by”. Precision can then be quantified by a dimensionless difference limen defined as the geometric standard deviation of X_{St} :

$$DL_G = \begin{cases} Co/St, & Co \geq St \\ St/Co, & Co < St \end{cases} = GS(X_{St}) . \tag{8}$$

Unlike the conventional difference limen, $DL_A = AS(X_{St})$, which must be divided by St to yield a relative measure of sensitivity, the proposed difference limen could be used directly as such a measure, because it is already such a ratio. For this reason it would be more correct to refer to it as a ratio limen, but by analogy to the conventional measure, we use the term geometric difference limen.

Note, however, that $DL_G \in \mathbb{R}_{>1}$, whereas we want Weber fractions to be elements of $\mathbb{R}_{>0}$, so that they are comparable to conventional Weber fractions. One might be tempted to define a novel Weber fraction as $W_G = |DL_G - 1| = |GS(X_{St}) - 1| = |GCV(X_{St})|$, which is the geometric coefficient of variation proposed by Kirkwood (1979). However, this would be inconsistent with the operations allowed on $\mathbb{R}_{>0}$. Instead, we propose to use the logarithm of DL_G ,

$$W_G = \ln(DL_G) = \text{dist}_{\mathbb{R}_{>0}}(Co, St) = \sigma , \tag{9}$$

because $\sigma \in \mathbb{R}_{>0}$ alone satisfies the properties required of the Weber fraction as a measure of relative sensitivity. Again, the log-ratio distance between Co and St is a dimensionless quantity with the auxiliary unit neper, which can be converted to decibels. Due to their common auxiliary units, CE_G and W_G can be directly compared. The proposed geometric measures are contrasted with the conventional measures in **Table I**, and we will use our experimental results to compare both types of measures in the sections that follow.

3.3. Dependence of Accuracy and Precision on Onset Duration, Sound Level, and Carrier of the Standard

The purpose of Experiment 1 was to examine how accuracy and precision of onset-duration matching depend on the onset duration, the sound level, and the carrier of the standard and to explore inter-listener variability. We compare the results when accuracy and precision are quantified by the conventional and the proposed measures. We find quantitative and qualitative differences which may even lead to opposite conclusions.

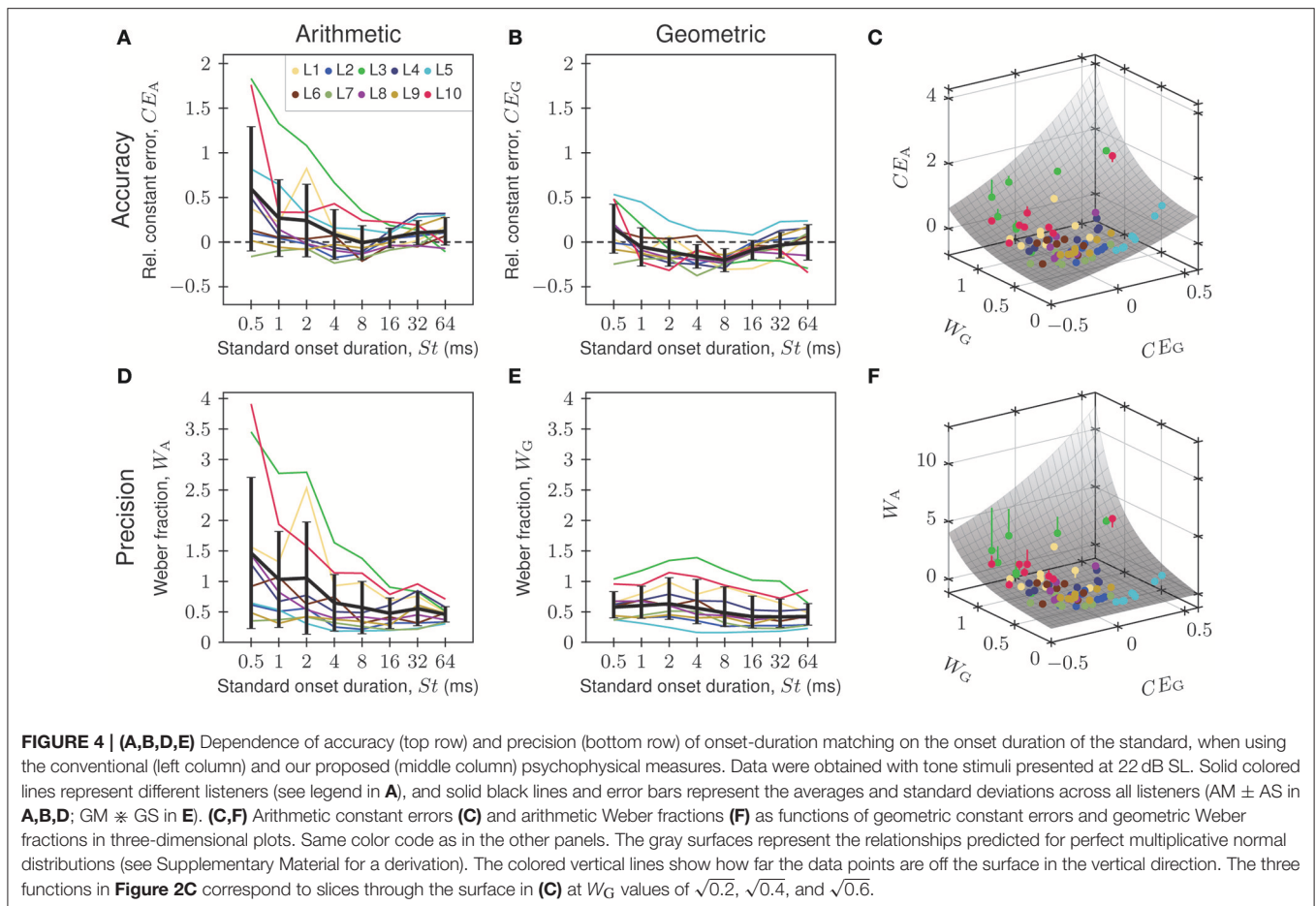
3.3.1. Inter-Listener Variability

Figure 4 shows, as an example, the constant error and the Weber fraction as a function of the onset duration of the standard for tones presented at 22 dB SL. Each colored line represents the data from a given listener ($n = 10$; see legend of **Figure 4A**), whereas the black solid lines and error bars represent the averages and standard deviations across listeners.

As shown in **Figure 4A**, the conventional measure of accuracy, the relative constant error CE_A , is usually positive, suggesting that listeners overestimate onset duration. The magnitude of CE_A tends to decrease with increasing onset duration of the standard. Furthermore, at short onset durations, values of CE_A differ widely among listeners (varying from -0.16 to 1.83) whereas at long onset durations, values are more similar (varying only from -0.11 to 0.32). This heteroscedasticity is reflected in the systematic and pronounced decrease of the arithmetic standard deviation of CE_A , as well as of the arithmetic mean of CE_A , with increasing onset duration of the standard. The null hypothesis of equal variance cannot be maintained (Bartlett test; $p < 10^{-6}$; $df = 7$).

In contrast, and as shown in **Figure 4B**, the proposed measure of accuracy, the relative constant error CE_G , is most often negative, revealing that listeners tend to underestimate onset duration. Only for listener L5 is CE_G consistently positive. The measure tends to be positive also for the shortest onset duration of the standard, an effect that may be attributed to the boundary effects imposed on the distribution of matches by the restricted range of comparison stimuli (see results of Experiment 3 below). Furthermore, the variation of CE_G across listeners is similar at all onset durations of the standard, and extreme values, as seen for CE_A at short onset durations, are not observed. This homoscedasticity of CE_G is reflected in similar arithmetic standard deviations of CE_G at all onset durations of the standard. The null hypothesis of equal variance of CE_G can be maintained (Bartlett test; $p = 0.0947$; $df = 7$). The variation of the arithmetic mean of CE_G with the onset duration of the standard is also smaller than the variation of the arithmetic mean of CE_A (cf. **Figures 4A,B**).

Similar observations were made for precision (**Figures 4D,E**). At short onset durations, values of the conventional measure of precision, the Weber fraction W_A , differ widely among listeners (from 0.35 to 3.91) whereas at long onset durations, values are more similar (varying only from 0.30 to 0.71). Again, this high degree of heteroscedasticity is reflected in the systematic and pronounced decrease of the arithmetic standard deviation of

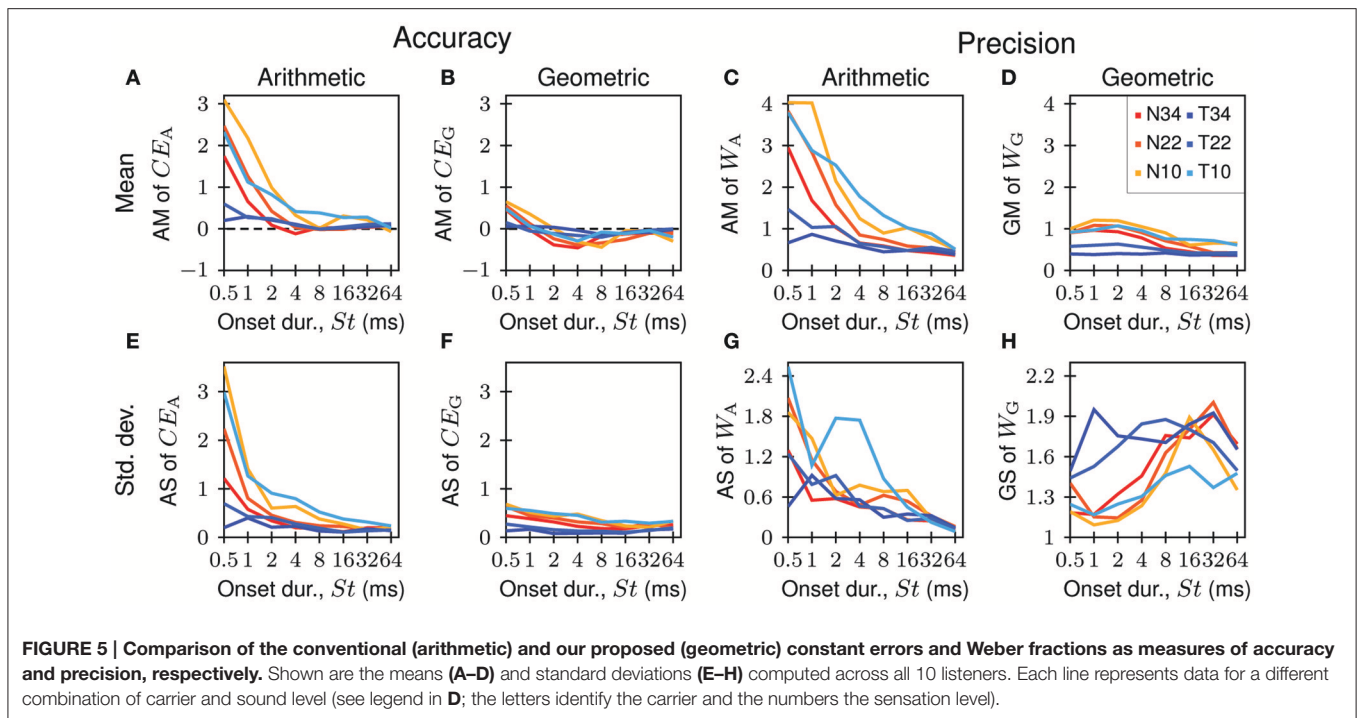


W_A with increasing onset duration of the standard, and the null hypothesis of equal variance of W_A cannot be maintained (Bartlett test; $p < 10^{-6}$; $df = 7$). The heteroscedasticity is also reflected in the monotonic decrease of the arithmetic mean of W_A , from about 1.5 for 0.5 ms to about 0.5 for 64 ms (**Figure 4D**). Also note that the arithmetic mean and arithmetic standard deviation do not characterize these data well. For example, only three listeners have larger, but seven listeners have lower, W_A values than the mean. Also, at several onset durations of the standard, the lower bound of the arithmetic standard deviation falls below the lowest W_A observed. These oddities arise because data that cannot be additive-normally distributed (W_A is always positive) have been characterized by arithmetic statistical measures.

In contrast, and as shown in **Figure 4E**, variation across listeners of the proposed measure of precision, the Weber fraction W_G , is similar at all onset durations of the standard. This homoscedasticity of W_G is reflected in similar geometric standard deviations of W_G at all onset durations of the standard. The null hypothesis of equal variance of W_G can be maintained (Bartlett test; $p = 0.5816$; $df = 7$). Furthermore, in individual listeners, the deviations of W_G from Weber's law are much smaller than for W_A , and extreme values, as seen for W_A at short onset durations, are not observed. Also, the geometric mean of W_G varies much

less with the onset duration of the standard (from about 0.4 to 0.6) than the arithmetic mean of W_A (cf. **Figures 4D,E**). Note that the geometric mean and the geometric standard deviation characterize these data well.

We also observed that the order of the listeners with respect to accuracy or precision can differ, depending on whether arithmetic or geometric measures are used. For instance, when CE_A is used, listener L3 appears to be the least accurate (**Figure 4A**), whereas when CE_G is used, listener L5 is the least accurate (**Figure 4B**). This can happen because the relationship between CE_A and CE_G is not order-preserving. Instead, CE_A depends on both CE_G and W_G . The gray surface in **Figure 4C** represents the relationship predicted for perfect multiplicative normal distributions (see Supplementary Material for a derivation). A large CE_A can result from a large W_G even when CE_G is small (e.g., as is the case for listener L3; see **Figure 4E**). Similarly, the relationship between W_A and W_G is also not order-preserving, because W_A also depends on both CE_G and W_G . The gray surface in **Figure 4F** represents the relationship predicted for perfect multiplicative normal distributions (see Supplementary Material for a derivation). The colored points in **Figures 4C,F** are derived from the data shown in **Figures 4A,B,D,E** (same color code). The colored vertical lines represent the vertical distances of the observed



CE_A and W_A from the surfaces. Most of them are short (not exceeding the symbols representing the data), such that the observed CE_A and W_A are generally close to those that would be obtained from perfect multiplicative normal distributions. These analyses corroborate our finding that the distributions of matches are well described by multiplicative normal distributions (cf. Figure 3).

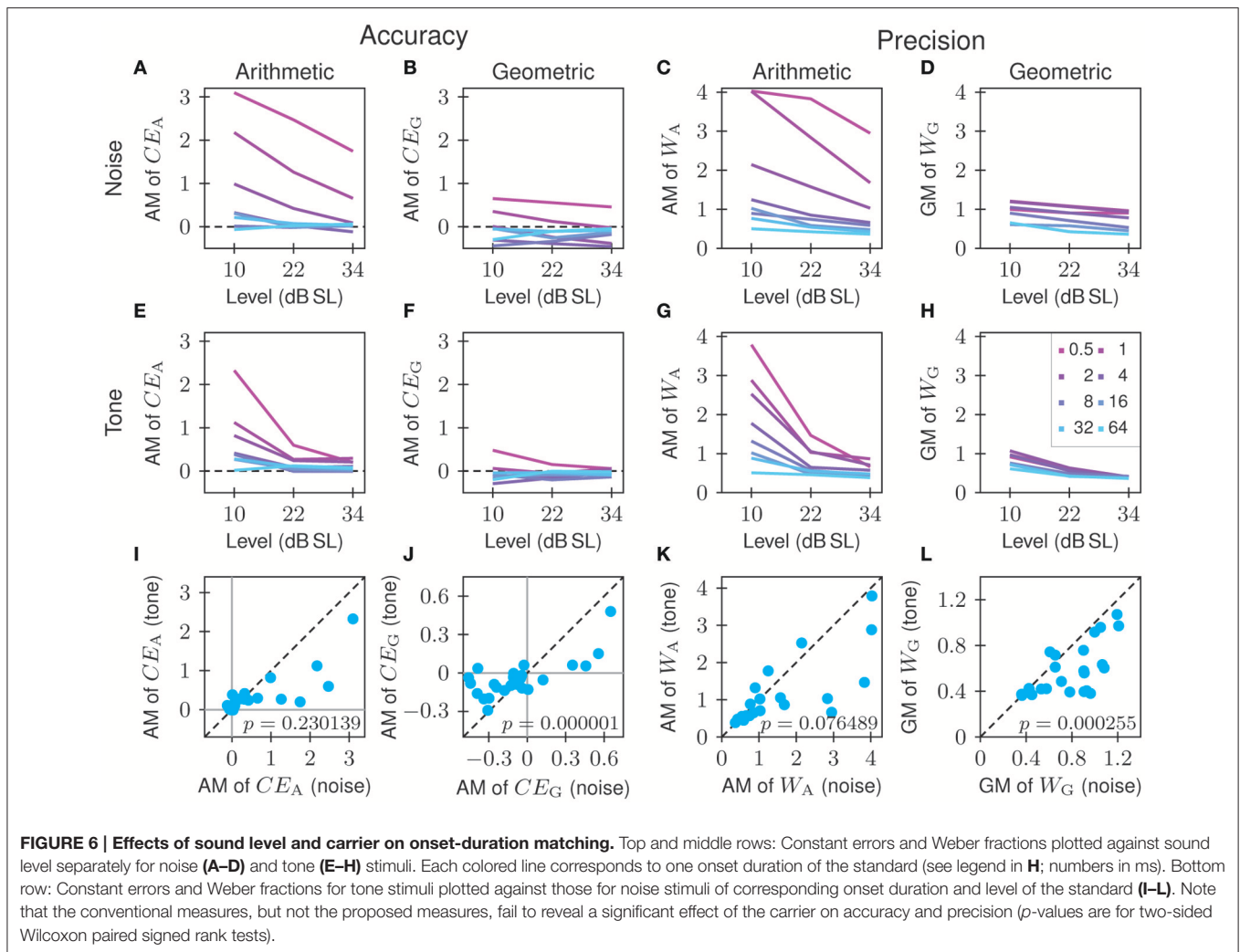
3.3.2. Grand-Mean Data

The results described so far, based on tones at 22 dB SL, apply in a qualitatively similar fashion to tones at the other sound levels and to the noise carrier. This is illustrated in Figure 5. Figure 5A shows the conventional measure of accuracy, the relative constant error CE_A , as a function of the onset duration of the standard. Each line represents the arithmetic mean of the CE_A across all 10 listeners for the tone or noise carrier at a given sound level (see legend in Figure 5D). The function for tones at 22 dB SL is identical to the mean function in Figure 4A. For both carriers and at all sound levels, the arithmetic mean of CE_A is predominantly positive and decreases with increasing onset duration of the standard in a manner that varies with sound level and carrier (see below). Figure 5E shows the corresponding arithmetic standard deviations of CE_A . For both carriers and all sound levels, the arithmetic standard deviation of CE_A decreases with increasing onset duration of the standard, reflecting the fact that the variance of CE_A across listeners is heterogeneous. The heteroscedasticity reflected in these curves decreases with increasing sound level for a given carrier. For a given sound level, the heteroscedasticity is more pronounced for the noise than for the tone carrier, but the null hypothesis of equal variance of CE_A cannot be maintained for any of the six combinations of

sound level and carrier (Bartlett tests; for all $p < 0.0025$ and $df = 7$).

In contrast, and as shown in Figure 5B, the arithmetic mean of CE_G is mostly negative, except at the two shortest onset durations of the standard. Again, this may be attributed to the boundary effects imposed on the distribution of matches by the restricted range of comparison stimuli (see results of Experiment 3 below). Furthermore, extreme values of the constant error such as those seen with CE_A are not observed for CE_G . Also, the variation of CE_G with the onset duration of the standard is much less pronounced than that of CE_A (cf. Figures 5A,B). The arithmetic standard deviation of CE_G changes little with the onset duration of the standard, reflecting a high degree of homoscedasticity (Figure 5F). Only for the noise carrier at 10 dB SL and 22 dB SL can the null hypothesis of equal variance of CE_G not be maintained (Bartlett test; $p = 0.0128$ and $p = 0.0064$, respectively; $df = 7$). For a given carrier, the arithmetic standard deviation of CE_G decreases with increasing sound level. For a given sound level, it is lower for the tone than for the noise carrier.

Figures 5C,D,G,H show the corresponding results for precision. Figure 5C shows the pronounced decrease of the arithmetic means of the conventional measure of precision, the Weber fraction W_A , with increasing onset duration of the standard for all sound levels and carriers. The function for 22 dB SL is the same as that shown in Figure 4D. This decrease is consistent with findings in the literature (cf. Figure 1) and implies that Weber's law does not hold under these conditions. The corresponding arithmetic standard deviations of W_A also decrease with increasing onset duration of the standard (Figure 5G), reflecting the high degree of heteroscedasticity of



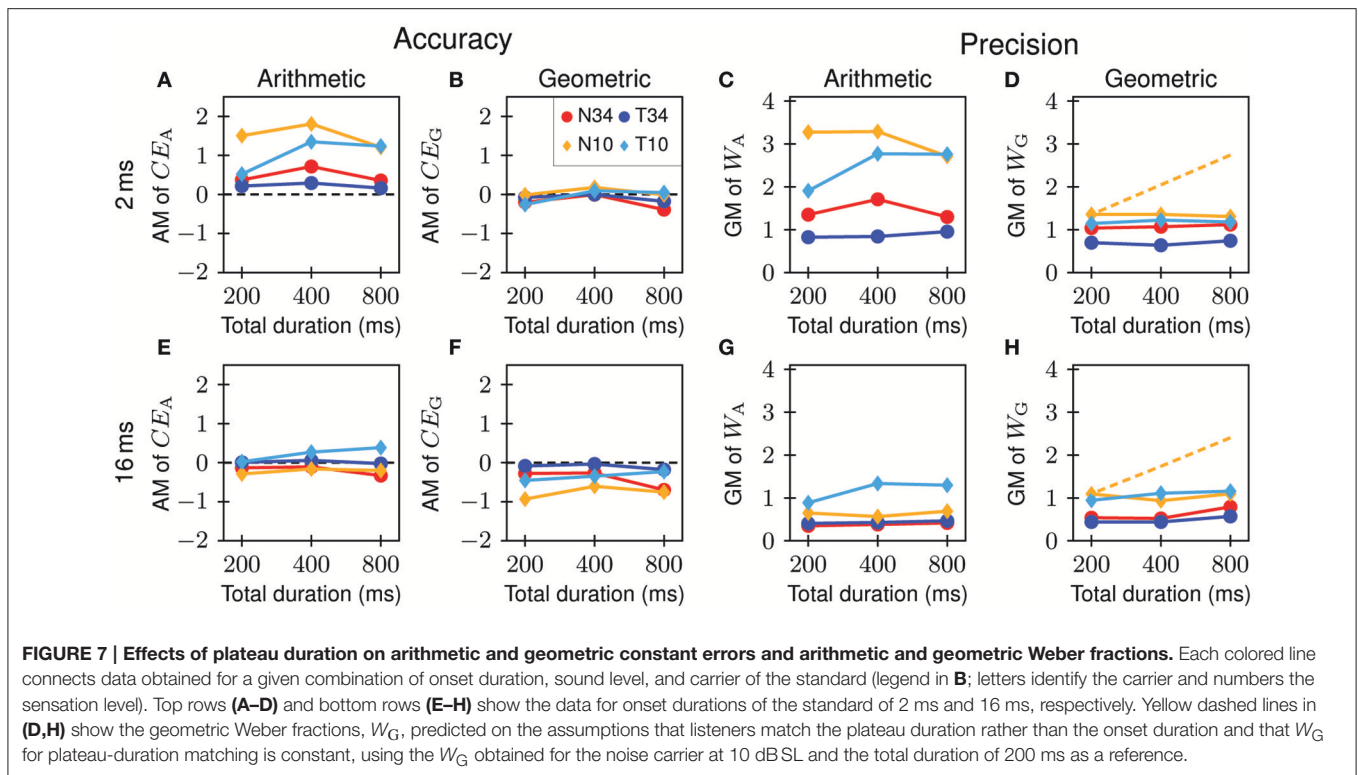
W_A . The null hypothesis of equal variance of W_A cannot be maintained for any of the six combinations of sound level and carrier (Bartlett tests; for all $p < 10^{-4}$ and $df = 7$).

In contrast, for both carriers and at all sound levels, the geometric mean of the proposed measure of precision, the Weber fraction W_G , changes little with onset duration of the standard (Figure 5D), although only for tones at 34 dB SL can Weber's law be said to apply (Friedman test; $p = 0.7643$). For tones at the lower two sound levels and for the noise carrier at all sound levels, the geometric mean of W_G initially increases with increasing onset duration of the standard from 0.5 to 1 or 2 ms before decreasing with further increases of the onset duration of the standard. The initial increase may again be attributed to the boundary effects imposed on the distribution of matches by the restricted range of comparison stimuli (see results of Experiment 3 below). Without such boundary effects, W_G may be expected to only decrease with increasing onset duration of the standard, but only slightly compared to W_A . The geometric standard deviation of W_G does not show the systematic decrease displayed by the arithmetic standard deviation of W_A (cf. Figures 5G,H). For tones at all sound levels and for the noise

carrier at 34 dB SL, the null hypothesis of equal variance of W_G can be maintained (Bartlett tests; all $p > 0.2265$ and $df = 7$). For the noise carrier at 10 dB SL and 22 dB SL, the null hypothesis of equal variance cannot be maintained (Bartlett test; $p = 0.0040$ and $p = 0.0361$, respectively; $df = 7$).

3.3.3. Effects of Sound Level and Carrier

The effects of sound level on accuracy and precision are summarized in the top and middle rows of Figure 6, in which the mean data from Figure 5 are replotted. The top row shows the data for the noise carrier and the middle row shows those for the tone carrier, plotted as functions of the sound level with each colored line corresponding to one of the eight onset durations of the standard (see legend in Figure 6H). Generally, accuracy and precision increase with increasing sound level, as reflected in the decrease of the magnitudes of the constant error (cf. Figures 6A,B,E,F) and the Weber fraction (cf. Figures 6C,D,G,H). The poorer the accuracy and precision for the lowest sound level, the stronger is the improvement with sound level. The



effects of sound level appear more pronounced and more dependent on the onset duration of the standard for the conventional (arithmetic) than for the proposed (geometric) measures.

The effect of the carrier on accuracy and precision are summarized in the bottom row of **Figure 6**, in which the mean of the measures for the tone carrier are plotted against those for the noise carrier of corresponding sound level and onset duration of the standard. Notably, when the conventional measures are used, the type of carrier has no significant effect on accuracy ($p = 0.2301$; Wilcoxon paired signed rank test; two-sided) or precision ($p = 0.0765$). In contrast, when the proposed measures are used, significant effects of the carrier on both accuracy ($p = 10^{-6}$) and precision ($p = 2.5 \cdot 10^{-4}$) are revealed.

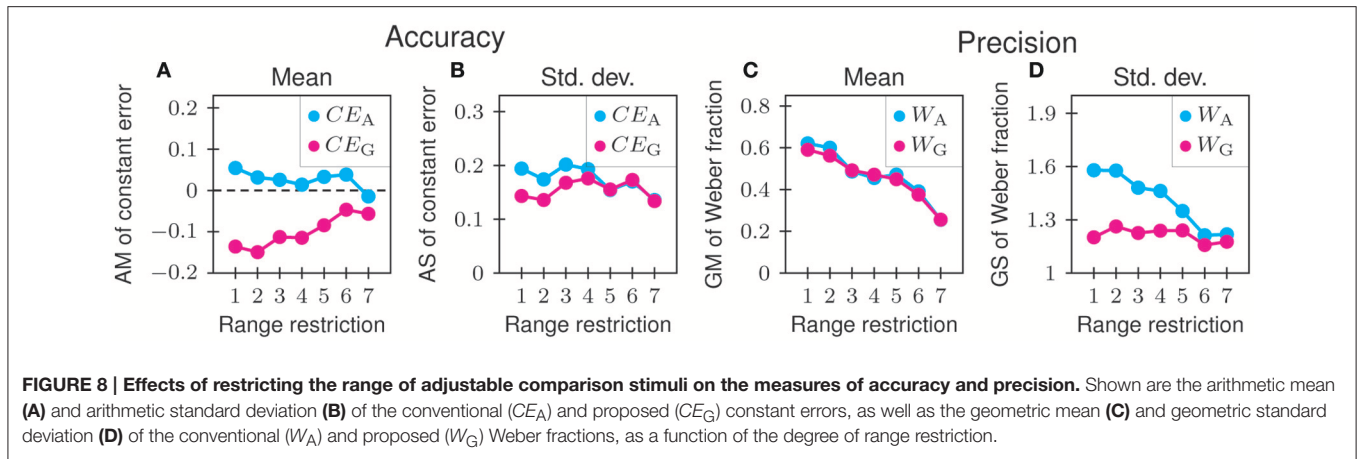
3.4. Accuracy and Precision of Onset-Duration Matching Do Not Systematically Depend on Plateau Duration

The purpose of Experiment 2 was to explore whether listeners use differences in the plateau duration, rather than in the onset duration, of the standard and comparison stimuli for matching. In Experiment 1, the total duration (400 ms) and the offset duration (50 ms) were fixed, such that plateau duration covaried with onset duration. If listeners actually matched plateau duration, and if plateau-duration matching approximately followed Weber's law, then the performance of the apparent onset-duration matching would improve by shortening, and worsen by prolonging, the plateau duration of the standard.

To test this idea, we employed standards with total durations of 400 ms (as in Experiment 1), 200 ms, and 800 ms. The offset duration was fixed at 50 ms (as in Experiment 1). Again, tone and noise carriers were used but with only two sound levels (10 dB SL and 34 dB SL) and only two onset-durations of the standard (2 and 16 ms).

Neither accuracy nor precision have a monotonic dependence on plateau duration (**Figure 7**). For the majority of combinations of sound level, carrier, and onset duration (7/8 for CE_A ; 5/8 for CE_G ; 8/8 for W_A and 6/8 for W_G), the null hypothesis of equal medians can be maintained (Friedman test; $p > 0.105$). For the remaining combinations, p -values were 0.0498 (4/6) or 0.0388 (2/6). In most cases, however, accuracy and precision changed nonmonotonically. In addition, precision changed much less than predicted if listeners matched plateau durations in accordance with Weber's law (represented by the yellow dashed lines in **Figures 7D,H**). It is therefore unlikely that listeners performed Experiments 1 and 2 by matching plateau durations. It is likely that listeners instead use onset cues, specifically differences in onset duration *per se* or associated spectral differences (see Discussion).

The results of Experiment 2, obtained with four listeners who did not take part in Experiment 1, also corroborate the differences in the arithmetic and geometric measures of accuracy and precision described in the previous section (**Figure 7**). They also corroborate the major findings of Experiment 1: accuracy and precision are higher for the tone than for the noise carrier, accuracy and precision increase with increasing sound level for both carriers, estimates of the constant error are more positive for short (2 ms) than for long (16 ms) onset durations of the



standard, and precision increases with increasing onset duration of the standard (Figure 7).

3.5. Effects of Restricting the Range of Comparison Stimuli

In Experiment 3, we examined the effects of restricting the range of adjustable comparison stimuli on the measures of accuracy and precision. The experiment was motivated by our observation in Experiments 1 and 2 that some of the distributions of matches for standards with short onset durations, low sound levels, and noise carriers, might have been subject to boundary effects due to restriction of the range of adjustable comparison stimuli. In Experiment 3, a single standard stimulus with an onset duration of 10.24 ms was used, and the range of adjustable comparison stimuli was increasingly restricted from values between 0.64 and 163.84 ms (first degree of restriction) to values between 6.64 and 15.79 ms (seventh degree of restriction; see Table 2).

Figure 8 shows the results of Experiment 3. The arithmetic mean of CE_A decreases with increasing range restriction from about 0.5 to about 0. The arithmetic mean of CE_G increases from about -0.15 and approaches 0 with increasing range restriction (Figure 8A). The absolute value of both measures of the constant error therefore diminishes as the range of adjustable comparison stimuli is restricted. The associated arithmetic standard deviation of CE_A also decreases with increasing range restriction, whereas that of CE_G shows no systematic change (Figure 8B). The geometric means of both W_A and W_G decrease monotonically with increasing range restriction, both from about 0.6 to 0.25 (Figure 8C). The decrease is expected because W_A and W_G are equal to 0 when only a single value of the comparison stimulus can be chosen. Precision therefore appears to improve with increasing range restriction. The geometric standard deviation of W_A decreases with increasing range restriction whereas that of W_G changes little over the seven degrees of range restriction we tested. Ultimately, however, the geometric standard deviations of W_A and W_G must approach 1 as the range of adjustable comparison stimuli is increasingly restricted. They are equal to 1 when only a single value of the comparison stimulus can be chosen. Boundary effects imposed on the distributions of

matches by the restricted range of comparison stimuli might therefore underlie the positive values of CE_G for short onset durations of the standard, as illustrated in Figure 5B, and the initial increase in W_G with increasing onset duration of the standard for some combinations of tone level and carrier, as illustrated in Figure 5D.

4. DISCUSSION

4.1. Convexity Bias, Illusions, and Novel Measures

We derived novel definitions of the psychophysical measures of the point of subjective equality, the constant error, the difference limen, and the Weber fraction for stimuli with positive-real magnitudes. Our aim was consistency: When two positive-real quantities are combined, then the result should also be positive real. This requirement is not always met when positive-real quantities are treated as if they were elements of the additive vector space of real numbers. The multiplicative Euclidean vector space of positive-real numbers endowed with the log-ratio distance function, however, fully satisfies this requirement. In a recent note, Graff (2014) promotes the use of the log-ratio distance function, arguing that relative differences between physical quantities are often more informative than absolute differences. Moreover, he makes the point that the determination of psychophysical laws first requires a clear-cut, objective expression of the physics of stimuli and responses. He shows that the log-ratio distance function exhibits all properties desired of a measure of relative differences, such as symmetry in selecting a reference, agreement between inverted units, and additivity. We build upon these ideas by linking the algebraic properties of the multiplicative Euclidean vector space with descriptive statistics. The geometric mean, unlike the arithmetic mean, is consistent with the structure of the multiplicative Euclidean space. Our Gedanken experiment (Figure 2) demonstrates that the conventional psychophysical measures possess a convexity bias of purely mathematical origin. As summarized in Figure 2C, the conventional arithmetic constant error would imply the existence of a positive bias or perceptual illusion (i.e., when

$CE_A > 0$), whose magnitude increases with the width of the distribution of matches, when no bias or illusion exists (i.e., when $CE_G = 0$). A true bias or illusion (i.e., when $CE_G \neq 0$) would be overestimated, missed, or underestimated, depending on its sign and magnitude and on the width of the distribution of matches. It is even possible, as in the present study, that $CE_A > 0$ while $CE_G < 0$ (see quadrant II in **Figure 2C**).

It may be criticized that the Gedanken experiment rests on the assumption that matches follow a multiplicative normal distribution and that the outcome may differ if the matches followed other distributions. However, the multiplicative normal distribution is the simplest model describing data that behave according to a multiplicative model (Limpert et al., 2001; Limpert and Stahel, 2011), and it is the maximum-entropy distribution on $\mathbb{R}_{>0}$ when only the mean and the variance are known (Park and Bera, 2009). Furthermore, the multiplicative normal distribution may be difficult to distinguish from other multiplicative distributions, such as the Pareto distribution, a power-law distribution (Newman, 2005; Clauset et al., 2009). We chose the multiplicative normal distribution because it allowed us to exploit the known relationships between its parameters and the statistical descriptive measures (Table 3 of the Supplementary Material) and to analytically determine the magnitude of the convexity bias. The particular distribution on $\mathbb{R}_{>0}$ is, however, not so relevant for the main argument: the convexity bias is a consequence of the AM-GM inequality, a special case of Jensen's inequality, and its existence is independent of the underlying distribution. The Gedanken experiment applies not only to the matching of onset durations, but also challenges matching studies in which the existence of perceptual illusions was inferred from, or their magnitudes were quantified by, the conventional constant error. This includes reports of the auditory time-stretching illusion (e.g., Sasaki et al., 2010) and also certain optical illusions such as the famous Müller-Lyer illusion (e.g., Restle and Decker, 1977; Bulatov et al., 1997; Bulatov and Bertulis, 1999; Woloszyn, 2010), the vertical-horizontal illusion (e.g., Higashiyama, 1996), and Titchener and Delboeuf illusions (e.g., Pressey, 1977). The impact of the convexity bias on measures of such perceptual illusions remains to be investigated. Another issue is how the theoretical considerations can be applied to other psychophysical methods, such as forced-choice procedures, where performance is quantified differently.

It may be argued that the problems we have identified above can be avoided if the conventional arithmetic analysis is performed on log-transformed data, as is often the case in the literature. Indeed, the geometric mean and standard deviation on which our proposed measures of accuracy and precision are based can be obtained by exp-transforming the arithmetic mean and standard deviation of the log-transformed random variable (see Table 1 of the Supplementary Material). However, log-transformation is often used only to make skewed distributions more symmetric (e.g., bell shaped), a requirement of many statistical tests. In hearing sciences, log-transformation of the stimulus amplitude (into decibels) serves mainly to compress the wide range of sound

pressures (> six orders of magnitude) or sound intensities (> twelve orders of magnitude) that are processed by auditory systems into a more manageable range of values and to respect the observation that equal increments in level correspond to roughly equal increments in sensation (Hassall and Zaveri 1988, p. 31; Moore 2013, p. 10; Pickles 2013, p. 3). When to log-transform data, when to back-transform measures obtained in log-space, or how to proceed with the back-transformed measures may often be only vaguely known. Because, historically, concepts in descriptive statistics have usually been developed based on the additive Euclidean vector space of real numbers, their mathematical relationships to the multiplicative Euclidean vector space of positive-real numbers may be little known. To avoid biases that are of purely mathematical origin, it is essential to respect that corresponding mathematical operations differ between the vector space of real numbers and that of positive-real numbers. For example, addition corresponds to multiplication, multiplication corresponds to exponentiation, and the Euclidean distance function $\text{dist}_{\mathbb{R}}(x, y) = |x - y|$ corresponds to the Euclidean distance function $\text{dist}_{\mathbb{R}_{>0}}(x, y) = |\ln(x) - \ln(y)|$. Our study may be the first in the field of perception science to demonstrate the mathematical necessity for geometric measures of performance when stimulus magnitudes can only be positive real. Our approach extends beyond simply log-transforming data and using arithmetic statistics.

4.2. Accuracy and Precision in Onset-Duration Matching

We reexamined onset-duration matching with respect to stimulus factors affecting the accuracy and precision of adjustments, as well as with respect to inter-listener variability measured with the conventional and the proposed psychophysical measures. We demonstrated that the multiplicative normal distribution usually fits the distributions of matches better than the additive normal distribution (**Figure 3**). This is in line with our theoretical considerations according to which matches of onset durations, or of other stimulus attributes that can attain only positive real magnitudes, should comply with a multiplicative rather than an additive model. We also showed that the superiority of the fit of the multiplicative normal distribution over that of the additive normal distribution depends on the width of the empirical distribution and on the number of matches. Unlike previous studies which are based on only few matches per listener and condition (e.g., van Heuven and van den Broecke, 1979; van den Broecke and van Heuven, 1983), the listeners in our study made many (up to 68) matches per condition. This is clearly sufficient to demonstrate the superiority of the multiplicative model and the use of geometric rather than additive descriptive statistical measures to characterize the distributions of matches and to quantify performance.

It may be argued that the superiority of the multiplicative model, demonstrated here (**Figure 3**), results from the logarithmic spacing of the adjustable onset durations of the comparison stimuli and that opposite results would have been

obtained with linear spacing. For the fitting procedure, however, it makes no difference how the samples are spaced. Nonetheless, it is conceivable that the spacing of the stimuli might affect the way listeners perform the tasks. For example, incrementing the onset duration by some amount above the standard constitutes a smaller relative change than decrementing it by the same amount below the standard. Such an asymmetry might bias the distribution of matches.

When comparing the behavior of the conventional and the proposed measures of accuracy and precision across listeners and conditions, we observed that the conventional measures display a high degree of heteroscedasticity and a strong dependence on the onset duration and the sound level of the standard. This would mean that the generalization gradients of performance differ markedly between listeners. In contrast, the proposed measures display a high degree of homoscedasticity and a weak dependence on the onset duration and the sound level of the standard (**Figures 4–6**), meaning that the generalization gradients of performance are similar for different listeners.

Furthermore, we found that the carrier of the standard has a significant influence on accuracy and precision of onset-duration matching, an effect detected when using the proposed geometric measures but missed with the conventional measures (**Figure 6**). We found that both accuracy and precision were better for the narrowband (tone) than the broadband (noise) carrier. This is consistent with the trend observed in the data of van Heuven and van den Broecke (1979) and van den Broecke and van Heuven (1983) and observed when comparing the findings of Kewley-Port and Pisoni (1984) with those of Smurzyński and Houtsma (1989) (see Introduction and **Figure 1B**). The better performance with narrowband than broadband carriers can be explained by assuming that listeners make use of additional spectral cues available when using the former. The spectral splatter induced by the onset of a tone increases with decreasing onset duration. These spectral differences might be exploited by the listeners, in addition to the temporal differences, when matching the onset durations of tones. With broadband carriers, the spectral differences are less pronounced or absent.

We found no systematic effect of total duration (and hence plateau duration) on accuracy or precision (**Figure 7**), suggesting that differences in plateau duration can be excluded as a dominant cue for onset-duration matching in our experiments. This is in line with a conclusion drawn by van Heuven and van den Broecke (1979). This leaves the temporal differences at the onset (and, as discussed above, their associated spectral differences) as potential cues. One crucial parameter may be the rate of rise of the amplitude at stimulus onset. This rate doubles when onset duration is halved and sound level is held constant. Under these conditions, accuracy and precision slightly decrease with increasing rate of rise, even with the geometric measures. On the other hand, the rate of rise of the amplitude at stimulus onset also increases when sound level is increased and onset duration is held constant. The rate of rise increases by a factor of four for a 12-dB increment in sound level. Under these conditions, however, accuracy and precision improve with increasing rate of rise. The rate of rise therefore cannot be

the only factor. The onset responses of cortical and subcortical neurons are very sensitive to the rate of rise of the stimulus amplitude, as observed in many studies that have varied onset duration or sound level (Hall and Feng, 1988; Phillips, 1988; Gooler and Feng, 1992; Heil, 1997a,b; Heil and Irvine, 1998; Lee et al., 2016), but onset responses are not a unique function of this rate. Based on the effects of varying onset duration, onset shape, and sound level on the onset responses of cat auditory cortical neurons, Heil and colleagues suggested that the first spikes of the responsive population of neurons track the onset envelope and the initial steady-state portion of the stimulus (for review, see e.g., Heil, 2003). They also concluded that each temporal envelope will evoke a unique spatiotemporal response pattern across the tonotopic and isofrequency axes of auditory cortical maps. Such unique patterns may be the basis for (or a physiological correlate of) the onset-matching capabilities and their dependence on onset duration, level, and carrier of the standard reported here.

Notably, the sign of the constant error in our data was affected by the type of measure used. When using the conventional measure, the relative CE_A values were typically positive (**Figures 4–8**), implying that listeners overestimated onset durations. Our analyses, however, show that in our data positive values of CE_A are likely the result of the convexity bias. The proposed measure of accuracy, the geometric constant error CE_G , was negative for most listeners and conditions. Values of CE_G tended to be positive only for the shortest onset duration of the standard, which might be due to a boundary effect imposed by the restricted range of comparison stimuli. The presence of predominantly negative values of CE_G may be due to presentation-order effects (Gescheider, 1997), given that the standard was always presented before the comparison stimulus. Whether this is a viable explanation could be assessed, if feasible, in future studies that also employ reversed and balanced (e.g., randomized) sequences of standard and comparison stimuli.

The results of Experiment 3 show that constant errors and Weber fractions converge toward zero when the range of values from which the comparison stimuli can be drawn is increasingly restricted (**Figure 8**). Thus, when the range of adjustable comparison stimuli is too narrow, the presence of systematic biases or illusions—no matter if positive or negative—will be missed, and precision will be overestimated. This supports a conclusion drawn by van Heuven and van den Broecke (1979), who argued that ceiling effects reduced the standard deviations computed from the distribution of matches to standards with the longest onset durations. A similar reduction of the standard deviations will occur for matches to standards with the shortest onset durations. Range effects therefore cannot be the origin of the rapid increase of conventional Weber fractions with decreasing onset duration at short onset durations of the standard.

4.3. Conclusions

Arithmetic descriptive statistics are appropriate for random variables over the additive Euclidean space \mathbb{R} , whereas geometric descriptive statistics should be used for random variables over the multiplicative Euclidean space $\mathbb{R}_{>0}$, in which relative differences are measured with the log-ratio distance function.

Our results imply that psychophysical studies should be reevaluated if arithmetic descriptive statistics were used to quantify performance when geometric statistics should have been used instead. This applies when the magnitudes of the stimuli can be only positive real, because then the measures of performance will be affected by the convexity bias. Unfortunately, it is not possible to simply compute the proposed geometric Weber fractions and geometric constant errors from the conventional arithmetic Weber fractions and arithmetic constant errors, because W_A and CE_A depend on both CE_G and W_G (lines in **Figure 2C**; prediction surfaces in **Figures 4C,F**). It will therefore be necessary to reanalyze the raw data. The size of the convexity bias will differ across studies because it depends on the width of the distribution, which will differ across conditions, tasks, and subjects.

AUTHOR CONTRIBUTIONS

BF conceived of the study, performed the experiments, developed the novel measures, analyzed the data, and wrote the manuscript.

REFERENCES

- BIPM (2006). *The International System of Units (SI), 8th Edn.* (Bureau International des Poids et Mesures). Available online at: <http://www.bipm.org/en/publications/si-brochure/>
- Bulatov, A., and Bertulis, A. (1999). Distortions of length perception. *Biol. Cybern.* 80, 185–193. doi: 10.1007/s004220050516
- Bulatov, A., Bertulis, A., and Mickiene, L. (1997). Geometrical illusions: study and modelling. *Biol. Cybern.* 77, 395–406. doi: 10.1007/s004220050399
- Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* 51, 661–703. doi: 10.1137/070710111
- Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. *Percept. Psychophys.* 31, 462–476. doi: 10.3758/BF03204856
- Cutting, J. E., and Rosner, B. S. (1974). Categories and boundaries in speech and music. *Percept. Psychophys.* 16, 564–570. doi: 10.3758/BF03198588
- Cutting, J. E., and Rosner, B. S. (1976). Discrimination functions predicted from categories in speech and music. *Percept. Psychophys.* 20, 87–88. doi: 10.3758/BF03198712
- Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *Q. J. Exp. Psychol. A* 33A, 185–207. doi: 10.1080/14640748108400785
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Hätel.
- Gerhardt, H. C., and Schul, J. (1999). A quantitative analysis of behavioral selectivity for pulse rise-time in the gray treefrog, *Hyla versicolor*. *J. Comp. Physiol. A* 185, 33–44. doi: 10.1007/s003590050363
- Gescheider, G. A. (1997). *Psychophysics. The Fundamentals, 3rd Edn.* New Jersey, NJ: Lawrence Erlbaum Associates, Inc.
- Gooler, D. M., and Feng, A. S. (1992). Temporal coding in the frog auditory midbrain: the influence of duration and rise–fall time on the processing of complex amplitude-modulated stimuli. *J. Neurophysiol.* 67, 1–22.
- Graff, C. (2014). Expressing relative differences (in percent) by the difference of natural logarithms. *J. Math. Psychol.* 60, 82–85. doi: 10.1016/j.jmp.2014.02.001
- Grey, J., and Gordon, J. (1978). Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.* 63, 1493–1500. doi: 10.1121/1.381843
- Guilford, J. P. (1954). *Psychometric Methods, 2nd Edn.* New York, NY; Toronto, ON; London: McGraw-Hill.
- Gupta, S. V. (2010). *Units of Measurement*. Heidelberg; Dordrecht; London; New York, NY: Springer Series in Materials Science. Springer.
- PH conceived of the study, analyzed the data, and wrote the manuscript.

FUNDING

This research is supported by the Deutsche Forschungsgemeinschaft (SFB/TRR 31 A6 to PH).

ACKNOWLEDGMENTS

We are grateful to Gabriele Schöps and Rike Schulz for running the experiments. We also thank Wojciech Kordecki, Artur Matysiak, Adam J. Peterson, and Werner A. Stahel for insightful discussions and comments on earlier versions of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.02013/full#supplementary-material>

- Hall, J. C., and Feng, A. S. (1988). Influence of envelope rise time on neural responses in the auditory system of anurans. *Hear. Res.* 36, 261–276. doi: 10.1016/0378-5955(88)90067-6
- Halpern, A. R., Zatorre, R. J., Bouffard, M., and Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* 42, 1281–1292. doi: 10.1016/j.neuropsychologia.2003.12.017
- Hary, J. M., and Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Percept. Psychol.* 32, 409–418. doi: 10.3758/BF03202770
- Hassall, J. R., and Zaveri, K. (1988). *Acoustic Noise Measurements*. Nærum: Brüel and Kjaer.
- Heil, P. (1997a). Auditory cortical onset responses revisited: I. First-spike timing. *J. Neurophysiol.* 77, 2616–2641.
- Heil, P. (1997b). Auditory cortical onset responses revisited: II. Response strength. *J. Neurophysiol.* 77, 2642–2660.
- Heil, P. (2003). Coding of temporal onset envelope in the auditory system. *Speech Commun.* 41, 123–134. doi: 10.1016/S0167-6393(02)00099-7
- Heil, P., and Irvine, D. R. (1998). The posterior field P of cat auditory cortex: coding of envelope transients. *Cereb. Cortex* 8, 125–141. doi: 10.1093/cercor/8.2.125
- Heil, P., Neubauer, H., Tiefenau, A., and von Specht, H. (2006). Comparison of absolute thresholds derived from an adaptive forced-choice procedure and from reaction probabilities and reaction times in a simple reaction time paradigm. *J. Assoc. Res. Otolaryngol.* 7, 279–298. doi: 10.1007/s10162-006-0042-y
- Hempstock, T. I., Bryan, M. E., and Webster, J. B. C. (1966). Free field threshold variance. *J. Sound Vibration* 4, 33–44. doi: 10.1016/0022-460X(66)90151-9
- Higashiyama, A. (1996). Horizontal and vertical distance perception: the discorded-orientation theory. *Percept. Psychophys.* 58, 259–270. doi: 10.3758/BF03211879
- Kalmylova, I. V., and Shakhshayev, S. A. (1988). Effects of tone intensity and frequency on discrimination of the sound stimulus transition process. *Neurosci. Behav. Phys.* 18, 378–383. doi: 10.1007/BF01193882
- Kewley-Port, D., and Pisoni, D. P. (1984). Identification and discrimination of rise time: is it categorical or noncategorical? *J. Acoust. Soc. Am.* 75, 1168–1176. doi: 10.1121/1.390766
- Kirkwood, T. B. L. (1979). Geometric means and measures of dispersion. *Biometrics* 35, 908–909.
- Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Percept. Psychophys.* 62, 1426–1439. doi: 10.3758/BF03212144

- Lee, C. M., Osman, A. F., Volgushev, M., Escabi, M. A., and Read, H. L. (2016). Neural spike-timing patterns vary with sound shape and periodicity in three auditory cortical fields. *J. Neurophysiol.* 115, 1886–1904. doi: 10.1152/jn.00784.2015
- Limpert, E., and Stahel, W. A. (2011). Problems with using the normal distribution—and ways to improve quality and efficiency of data analysis. *PLoS ONE* 6:e21403. doi: 10.1371/journal.pone.0021403
- Limpert, E., Stahel, W. A., and Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. *BioScience* 51, 341–352. doi: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2
- Mapp, A. P., and Ono, H. (2006). *Precision and Accuracy with Three Psychophysical Methods (English Version)*. Available online at: <http://www.yorku.ca/psycho/index.html>
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol. Res.* 58, 177–192. doi: 10.1007/BF00419633
- Moore, B. C. H. (2013). *An Introduction to the Psychology of Hearing, 6th Edn.* Leiden; Boston, MA: Brill.
- Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* 46, 323–351. doi: 10.1080/00107510500052444
- Paquette, C., and Peretz, I. (1997). Role of familiarity in auditory discrimination of musical instrument: a laterality study. *Cortex* 33, 689–696. doi: 10.1016/S0010-9452(08)70726-3
- Park, S. Y., and Bera, A. K. (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *J. Econometr.* 150, 219–230. doi: 10.1016/j.jeconom.2008.12.014
- Phillips, D. P. (1988). Effect of tone-pulse rise time on rate-level functions of cat auditory cortex neurons: excitatory and inhibitory processes shaping responses to tone onset. *J. Neurophysiol.* 59, 1524–1539.
- Pickles, J. O. (2013). *An Introduction to the Physiology of Hearing, 4th Edn.* Leiden; Boston, MA: Brill.
- Pollack, I. (1963). Onset discrimination for white noise. *J. Acoust. Soc. Am.* 35, 607–609. doi: 10.1121/1.1918547
- Pressey, A. W. (1977). Measuring the Titchener circles and Delboeuf illusions with the method of adjustment. *Bull. Psychon. Soc.* 10, 118–120. doi: 10.3758/BF03329298
- Ramajanickam, M. (2002). *Modern Psychophysical and Scaling Methods and Experimentation*. New Delhi: Concept Publishing.
- Restle, F., and Decker, J. (1977). Size of the Mueller-Lyer illusion as a function of its dimensions: theory and data. *Percept. Psychophys.* 21, 489–503. doi: 10.3758/BF03198729
- Rosen, S. M., and Howell, P. (1981). Plucks and bows are not categorically perceived. *Percept. Psychophys.* 30, 156–168. doi: 10.3758/BF03204474
- Rosen, S. M., and Howell, P. (1983). Sinusoidal plucks and bows are not categorically perceived, either. *Percept. Psychophys.* 34, 233–236. doi: 10.3758/BF03202950
- Sasaki, T., Nakajima, Y., ten Hoopen, G., van Buuringen, E., Massier, B., Kojo, T., et al. (2010). Time stretching: illusory lengthening of filled auditory durations. *Atten. Percept. Psychophys.* 72, 1404–1421. doi: 10.3758/APP.72.5.1404
- Smurzyński, J. (1985). Noncategorical identification of rise time. *Percept. Psychophys.* 38, 540–542. doi: 10.3758/BF03207063
- Smurzyński, J., and Houtsma, A. J. (1989). Auditory discrimination of tone-pulse onsets. *Percept. Psychophys.* 45, 2–9. doi: 10.3758/BF03208025
- Tenney, J. C. (1962). Discriminability of differences in the rise time of a tone. *J. Acoust. Soc. Am.* A34:739. doi: 10.1121/1.1937296
- van den Broecke, M. P., and van Heuven, V. J. (1983). Effect and artifact in the auditory discrimination of rise and decay time: Speech and nonspeech. *Percept. Psychophys.* 33, 305–313. doi: 10.3758/BF03205877
- van Heuven, V. J., and van den Broecke, M. P. R. (1979). Auditory discrimination of rise and decay times in tone and noise bursts. *J. Acoust. Soc. Am.* 66, 1308–1315. doi: 10.1121/1.383551
- von Helversen, D. (1993). 'Absolute steepness' of ramps as an essential cue for auditory pattern recognition by a grasshopper (*Orthoptera; Acrididae; Chorthippus biguttulus L.*). *J. Comp. Physiol. A* 172, 633–639. doi: 10.1007/BF00213685
- Woloszyn, M. R. (2010). Contrasting three popular explanations for the Muller-Lyer illusion. *Curr. Res. Psychol.* 1, 102–107. doi: 10.3844/crsp.2010.102.107

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Friedrich and Heil. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.