# Assessing the Validity of Subjective Reports in the Auditory Streaming Paradigm

Dávid Farkas[a)]

*Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary*

Susan L. Denham

*Cognition Institute and School of Psychology, University of Plymouth, Plymouth, UK*

Alexandra Bendixen

*School of Natural Sciences, Chemnitz University of Technology, Chemnitz, Germany*

István Winkler

*Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary*

[a)] Author to whom correspondence should be addressed. Electronic mail: farkas.david@ttk.mta.hu

Abstract

While subjective reports provide a direct measure of perception, their validity is not self-evident. Here, we tested three possible biasing effects on perceptual reports in the auditory streaming paradigm: errors due to imperfect understanding of the instructions, voluntary perceptual biasing, and susceptibility to implicit expectations. 1) Analysis of the responses to catch trials separately promoting each of the possible percepts allowed us to exclude participants who likely have not fully understood the instructions. 2) Explicit biasing instructions led to markedly different behavior than the conventional neutral-instruction condition, suggesting that listeners did not voluntarily bias their perception in a systematic way under the neutral instructions. Comparison with a random response condition further supported this conclusion. 3) No significant relationship was found between social desirability, a scale-based measure of susceptibility to implicit social expectations, and any of the perceptual measures extracted from the subjective reports. This suggests that listeners did not significantly bias their perceptual reports due to possible implicit expectations present in the experimental context. In sum, these results suggest that valid perceptual data can be obtained from subjective reports in the auditory streaming paradigm.

PACS numbers: 43.66.-x, 43.66.Lj

# I. INTRODUCTION

The term bi-/multistable perception refers to the sensory phenomenon of conscious perceptual awareness switching stochastically and to some extent uncontrollably between alternative interpretations of an unchanging stimulus (for a review, see Schwartz, Grimault, Hupé, Moore, & Pressnitzer, 2012). Perceptual bi-/multistability are often studied by asking participants to continuously report their perception of the stimulus. Such subjective reports have advantages over 'objective' tests that measure performance in some task assumed to be facilitated when experiencing one of the alternative percepts (for discussion see Bendixen, 2014). Subjective reports provide a direct measure of the listener's perception, as opposed to the indirect measures inferred from task performance in 'objective' tests. They also reduce the effects of the listener's task-related strategies on the assessment of perception (i.e., listeners may bias their perception for better task performance or use some strategy not accounted for by the experimenter when inferring perception from task performance). Subjective reports can also be recorded continuously whereas most 'objective' task measures provide information only at discrete time points. Therefore, subjective reports allow comparisons between brain activity observable in different perceptual states (within the same listener and evoked by the same stimulation) without the confounding effects of success vs. failure in some task. However, the validity of subjective reports requires careful examination, because the experimenter cannot directly verify the truthfulness of these reports. In general, one has no reason to believe that volunteers do not faithfully report their perceptions. Nevertheless, unintended biases may be evoked by the experimental environment, such as the formulation of the instructions, the nonverbal behavior of the experimenter, etc. These kinds of biases could especially affect the reports of listeners who are sensitive to social expectations. The current study assesses the effects of such biases using the auditory streaming paradigm (an auditory multistable stimulus serving as the primary test-bed of auditory stream segregation, cf. van Noorden, 1975), in four different ways: 1) by analyzing the responses to catch trials, which promote the perception of one of the alternatives, 2) by assessing the effects of explicitly instructing listeners to bias their perception, 3) by examining the effects of listeners' sensitivity to social expectations (termed the social desirability bias; Paulhus, 1991) on the variables extracted from the subjective reports, and 4) by comparing participants' deliberately random responses to their responses when asked to report their perceptions.[1]

Separating concurrent and/or interleaved sounds (termed ‘auditory stream segregation’ by Bregman, 1990) is an essential function of the human auditory system. In the auditory streaming paradigm, a sequence made up of a repeating ABA- sound pattern (where A and B denote two different sounds and ‘-’ stands for a silent interval with the same duration as the two sounds; FIG.1, topmost panel; van Noorden, 1975), has been widely used to study the processes underlying stream segregation. Such sequences can be perceived in a number of different ways, the most common being: 1) all sounds are grouped together into a single coherent sequence that gives the impression of a galloping rhythm (termed the *integrated percept*; FIG.1, second panel from the top) and 2) the A and B sounds are grouped separately, resulting in the perception of two parallel isochronous streams of sound (A- and -B--) with one appearing in the foreground and the other in the background (termed the *segregated percept*; FIG.1, third panel from the top). When given the option, listeners are also able to report other repeating patterns, which have been termed the *both* response in previous reports (FIG.1, bottom panel, Denham, Gyimesi, Stefanics, & Winkler, 2010). In the current study we will refer to these patterns as *combined*, because they combine features of the integrated and the segregated percepts.

In the auditory streaming paradigm, perception is primarily influenced by the distinctiveness/similarity of the A and B sounds (e.g. differences in pitch, location, timbre, etc.) and the presentation rate (for a review see Moore & Gockel, 2012). However, given sufficient time, perception has been shown to inevitably switch back and forth between the alternative percepts (Bendixen, Denham, Gyimesi, & Winkler, 2010; Deike, Heil, Böckmann-Barthel, & Brechmann, 2012; Denham & Winkler, 2006; Gutschalk et al., 2005; Pressnitzer & Hupé, 2006) even for stimulus parameter combinations that have been assumed to strongly promote one of the perceptual alternatives (Denham, Gyimesi, Stefanics, & Winkler, 2013). Although switching between alternative percepts has been found to occur inevitably (Kashino, Kondo, & Okada, 2008; Pressnitzer & Hupé, 2006), at least for a large part of the parameter space, listeners can voluntarily bias their perception towards one or another alternative percept (van Noorden, 1975; for a review of top-down effects on bi-stable perception, see Snyder, Gregg, Weintraub, & Alain, 2012). In order to study the validity of subjective reports in the auditory streaming paradigm, we developed a simplified version of our previous protocol for assessing individual switching patterns in the auditory streaming paradigm (Denham et al., 2014).

Four possible sources of bias in the subjective perceptual reports were examined in the current study: 1) imperfect understanding of the instructions by the listener, such as confusion about how to categorize and respond to the various perceived patterns; 2) the listener voluntarily biasing their perception; 3) susceptibility to implicit expectations attributed to the experimenter. That is, beyond possible confusion/errors in reporting their percepts (point 1), listeners may bias their perception explicitly then report it truthfully (point 2) or bias either their perception or the report of it due to perceived implicit expectations (point 3). In a follow-up experiment, we also tested the possibility that, knowing that the experimenter cannot directly verify their responses, listeners have altogether disregarded the sounds and randomly pressed the response keys.

Imperfect understanding of the instructions can be tested by including catch trials in the experimental design. Catch trials for the auditory streaming paradigm were created by using parameter combinations that strongly promote one of the perceptual alternatives. Participants who did not reliably detect the patterns on which they were trained can be rejected based on their catch trial matching (CTM) accuracy. However, the rejection criterion is largely arbitrary. Therefore, we assessed the effects of the rejection criterion on intra-individual consistency in the subjective perceptual reports. The underlying assumption was that participants who did not fully understand, remember, or follow the instructions would not report their perception in a consistent manner.

For testing the effects of voluntary perceptual biasing, listeners were instructed to either hold each percept for as long as they could or to switch to another percept as soon as they could. Both of these instructions have been found to affect the rate of switching between the perceptual alternatives for visual bi-stable configurations (for holding percepts as long as possible, see e.g., Toppino, 2003; for switching as often as possible, see e.g., van Ee, van Dam, & Brouwer, 2005). In the auditory modality, Pressnitzer and Hupé (2006) tested the effects of instructing participants to hold on to either the integrated or the segregated percept for as long as they could. They found that switching between alternatives occurred even in this case. They also found that the average duration of the periods in which the designated sound organization was perceived (termed average phase duration) as well as the proportion of perceiving this organization increased compared to the neutral-instruction condition. Thus, listeners can voluntarily bias the length of the time they experience the alternative perceptual states, but their control over their perception is also limited. We expected the number of switches in the neutral-instruction blocks to fall between those obtained when listeners are instructed to a) hold their percept as long as they can and b) switch between percepts as often as they can. Thus these data should place boundaries on the possible effects of voluntary perceptual biasing by listeners in the neutral-instruction conditions. Another possibility for voluntary biasing was that listeners assumed that the perceptual alternatives should be experienced with equal probability. This possibility was tested against the observed responses.

Listeners reporting their perception in a biased way due to perceived implicit expectations from the experimenters is an example of the *social desirability bias* (SDB) phenomenon, which refers to the '*systematic error in self-report measures resulting from the desire of respondents to avoid embarrassment and project a favorable image to others*' (Fisher, 1993, p. 303). One solution for preventing such effects is to use a double-blind experimental design with carefully assessed instructions. However, collecting subjective reports in the auditory streaming paradigm requires a balance between preassigned wording and individually tailored explanations for participants to understand their task, which naïve experimenters may not be able to provide. Here we provide an estimate of the social desirability effect on the measures extracted from subjective reports obtained in the auditory streaming paradigm by assessing the listeners' susceptibility to implicit expectations.

Two dimensions of the SDB have been identified (Paulhus, 1984, 1991; Zerbe & Paulhus, 1987). The first one is *self-deceptive positivity* (SDP), which is a tendency to protect the self from threats by positively biased responding. This is an honest bias, because the respondent believes that his or her answers are truthful. The second dimension is *impression management* (IM), which refers to the tendency to present oneself as favorably as possible to others. In other words, responses are consciously and deliberately altered to present a good image of the respondent. SDP is considered to be an invariant personality trait, while IM is only active when other people are present in a situation (Paulhus, 1984, 1991; Zerbe & Paulhus, 1987). Psychological studies using questionnaires have long been used to screen and control for participants based on their SDB score (e.g., Fisher & Katz, 2000), either by excluding participants who have a high tendency for socially desirable responding, or by taking the phenomenon into account during the statistical analysis (King & Bruner, 2000).

For assessing the effects of implicit expectations attributed to the experimenter, we assumed that listeners with a higher tendency for SDB would be more susceptible to possible implicit expectations and thus the bias would induce some systematic tendencies in the variables extracted from the subjective reports as a function of the level of SDB. Therefore, using correlation analysis and regression models, we assessed the effects of SDB on the variables extracted from the subjective reports. In addition, to estimate the effects of SDB at the group level, we systematically rejected increasing numbers of listeners with the highest SDB from the group and compared the results across the remaining (overlapping) groups (i.e., simulating the effects of excluding participants on the basis of high SDB).[2]

Finally, in a follow-up study, we wished to verify that the responses of the listeners' were in fact based on their perception of the sounds by comparing the switching patterns produced under the perceptual report instructions with those produced when the same participants were instructed to produce random responses. The distribution of perceptual reports is expected to be either gamma (if these perceptual decisions result from a combination of discrete stochastic processes; Zhou, Gao, White, Merk, & Yao, 2004) or lognormal (if they are the product of independent stochastic processes; Murata, Matsui, Miyauchi, Kakita, & Yanagida, 2003). The distribution of button presses produced when participants were instructed to press randomly has been found to differ from either of the above distributions (Leopold, 1997). The distribution of perceptual reports obtained in the auditory streaming paradigm has been found to be consistent with a lognormal distribution (Pressnitzer & Hupé, 2006). However, in this case perceptual reports were not compared with random button presses.

In summary, the aim of the current study is to assess the validity of subjective reports in the auditory streaming paradigm 1) through the screening of participants by catch trials appended to the test blocks; 2) by demonstrating the effects of voluntary control on perception in the auditory streaming paradigm in order to provide bounds on the variables extracted from subjective reports; 3) by assessing the effects of social desirability bias on the variables extracted from subjective reports obtained in the auditory streaming paradigm, and 4) by comparing responses obtained as perceptual reports with those produced with random-press instructions.

## II. METHODS

### A. Participants

Seven volunteers (5 males), aged between 20 and 26 years ($M_{\text{age}}$ = 22.14, $SD_{\text{age}}$ = 2.04) participated in the pilot study. Fifty-three healthy volunteers (77.4% female), aged between 18 and 42 years ($M_{\text{age}}$ = 22.38, $SD_{\text{age}}$ = 4.04) participated in the main experiment. One participant's data were excluded from this experiment, because she reported hearing the integrated percept throughout the test blocks. While this is a valid response, the participant is an extreme outlier within the group, one whose perceptual tendencies cannot be correctly represented in the statistical analyses. Thus, the sample analyzed was based on fifty-two participants (76.9% female), aged between 18 and 42 years ($M_{\text{age}}$ = 22.44, $SD_{\text{age}}$ = 4.05). Fourteen volunteers (4 males), aged between 19 and 26 years ($M_{age}$ = 22.57, $SD_{age}$ = 2.53), participated in the follow-up experiment. All participants reported having normal hearing. None of the participants were taking any medication affecting the central nervous system or took part in more than one experiment. Participants gave written informed consent after the experimental procedures had been explained to them. The study was approved by the institutional review board of the Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences. Participants either received modest financial compensation or extra credits in a university course for their participation.

**B. Stimuli and Apparatus**

Sinusoidal tones of 75 milliseconds (ms) duration (including 10ms rise and fall times) with an intensity of 45 dB sensation level (hearing threshold measured individually for each participant with a staircase procedure using the experimental sounds) were arranged according to the auditory streaming paradigm (a cyclically repeating 'ABA-' pattern). The frequency difference was 4 semitones with the 'A' tone's frequency set at 400 Hz and 'B' tone frequency at 504 Hz. The stimulus onset asynchrony (SOA, onset-to-onset time interval) was 150 ms. Participants were presented with 4-minute-long ABA- tone sequences. An additional ca. 40 second catch-trial segment (see the Tasks and Procedures section) was appended without a break to the end of each block. Tones were delivered through Sennheiser HD600 headphones (Sennheiser electronic GmbH & Co. KG) by an IBM PC computer using the Cogent 2000 stimulus presentation software under MATLAB (MathWorks Inc.).

**C. Tasks and Procedures**

Listeners were instructed to mark their perception continuously in terms of four possible categories (see FIG.1): a) integrated (ABA-; depressing one of the two response keys), b) segregated (A-A-/-B-- or -B--/A-A-, depending which stream was heard in the foreground; depressing the other response key), c) combined (AB--/--A-, -BA-/A---, see Denham et al., 2014 for details; simultaneously depressing both response keys), and d) none (none of the above patterns being perceived; releasing both response keys). Participants received both descriptions and training for the different perceptual categories. The description of the integrated percept emphasized hearing all tones as part of a single repeating pattern with a galloping rhythm. The description of the segregated percept emphasized hearing two isochronous sound streams in parallel, one in the foreground, the other in the background, each with a uniform rate (one slower, the other faster). The description of the combined percept emphasized the perception of two parallel streams of sound, at least one of which included a repeating pattern composed of both high and low tones. Finally, the 'none' choice allowed listeners to indicate that they did not hear any repeating pattern or could not decide between the patterns previously described to them. The left and right arrow keys of a standard Hungarian IBM PC keyboard were used as response keys; their role was balanced across participants. Participants were instructed to keep one or both keys depressed for as long as they continued hearing the corresponding pattern, but to switch to another combination as soon as their perception changed. Instructions emphasized that their perception might change and that there are no correct or incorrect answers in the task. We also asked them not to voluntarily try perceiving the sounds in one or another way; this set of instructions will be referred to as the *neutral instructions*. The state of the response keys was continuously recorded at a nominal sampling rate of 10 Hz using MATLAB and the Cogent 2000 software.

Participants were trained to indicate their perception in terms of the above listed four categories without hesitation. Training started with the participant listening to six one-minute long auditory streaming sequences, each promoting the perception of one of the alternatives shown in FIG.1. The integrated percept was introduced by using a smaller frequency difference between the ‘A’ and ‘B’ tones (1ST; 400 and 426 Hz, respectively) than the parameters chosen for the experiment, while the segregated percept was initially demonstrated by using a larger frequency difference (400 and 890 Hz, respectively). Subsequently, the two segregated and three of the combined percepts in FIG.1 were demonstrated by emphasizing the corresponding repeating tone pattern within an auditory streaming sequence with the parameters used in the experiment. This was done by attenuating by 18 dB those tones which were not part of the intended foreground stream (thus promoting perception of the normal-intensity tones as a single coherent sequence). After the response key assignment and the ‘none’ choice was explained, training continued in blocks of six sequences separated by short silent intervals. The first sequence was 1 minute long and its parameters were identical to those used in the experiment. This was followed by five sequences of 6-9 s, each, one sequence for each of the examples presented initially to the participant, except for the large frequency-difference version of the segregated percept demonstration (to avoid confusion regarding the segregated percept). The order of the five catch trial sequences (low-frequency difference and the 4 attenuated-intensity examples) was randomized separately for each training block. The training was completed when the participant made the intended response during at least 80% of the presentation time of the examples or once 20 training blocks had been delivered. During the training blocks, the experimenter gave feedback and further help as needed. No participant was rejected on the basis of the accuracy of their responses within the training blocks.

The ca. 40-second catch-trial (30-45 s) appended to the end of each four-minute long block in the experiment consisted of five 6-9 s long segments of the example patterns used in the training blocks. The order of the example patterns was randomized across participants, and segments with each pattern appeared exactly once in a catch-trial.

In the pilot experiment, once their training had been completed, participants received three blocks with the neutral instructions, after which they completed the additional tasks and questionnaires[3]. Finally, they received another three blocks, again with the neutral instructions. Breaks were included when switching tasks and between blocks as needed. The session lasted for about 150 minutes.

In the main experiment, the auditory streaming segment consisted of the training part followed by 11 experimental blocks. For the first five blocks, participants reported their perception according to the neutral instructions. For three of the remaining blocks, participants were instructed to hold on to each percept as long as they could (Hold condition) and for the other three blocks, participants were instructed to switch to another percept as fast as they could (Switch condition). In both of these conditions, participants were instructed to continue to report their perception faithfully (i.e., they were asked to bias their perception, not their reporting behavior). The order of these two biased conditions was counterbalanced between participants. Breaks were included when switching tasks and between blocks as needed. The session lasted for about 180 minutes.

Before and after the auditory streaming tasks, participants completed several other tasks and questionnaires[3]; only one of which is reported here. The Balanced Inventory of Desirable Responses (BIDR, Paulhus, 1984, 1991) measures two dimensions, *self-deceptive positivity* (SDP) and *impression management* (IM). Both dimensions are measured by 20 items.

Respondents indicate their agreement with items on a seven-point Likert scale ranging from 'Not true' to 'Very true'. Based on the results of Stöber, Dette, and Musch (2002) scores were calculated by summing the item scores separately for each dimension.

In the follow-up experiment, participants first performed the random conditions. They were instructed to switch between the three combinations of button states described for the pilot and the main experiments (i.e., one or the other or both buttons depressed) randomly (Random-Neutral condition). Next they were instructed to either switch more often (Random-Switch condition), or less often (Random-Hold condition) between the button states than they did in the first condition. During these three blocks (4 minutes, each), participants listened to the same sound sequences as were presented in the main experiment. They were informed that the sounds were not related to their task. After the random conditions, participants completed the training for reporting their perception in the auditory streaming paradigm. The training was identical to that described for the main experiment, except that the duration of the last training block was identical to that of the experimental blocks (i.e. 4 minutes). After the training, participants completed four blocks with instructions to report their perception as was described for the main experiment: the first two blocks were performed under the neutral instructions (Auditory-Neutral condition), then one block under the Hold (Auditory-Hold condition) and another under the Switch instructions (Auditory-Switch condition). The instructions were identical to those described for the main experiment. The order of the two biased-condition blocks was counterbalanced across participants.

## D. Data Preprocessing

As a first preprocessing step, perceptual phases (continuous periods during which the same perception is reported) that were shorter than 300 ms were discarded because these were assumed to result from small inaccuracies when switching between key combinations (i.e., when the response mapping requires changing two keys simultaneously, synchronization between the two movements is usually imperfect, resulting in a short phase with a third key combination; see Moreno-Bote, Shpiro, Rinzel, & Rubin, 2010). The data removed by this preprocessing step amounted to 0.5% of the total data in the main experiment.

Transition matrices were constructed from the resulting records using the method described by Denham et al. (2012). In the current study, each transition matrix is a 4x4 matrix with elements that represent the probability of switching from one percept to another (i.e. four possible response states: three possible perceptual categories and the 'none' choice), with the percept of origin corresponding to the column and the destination percept to the row of the matrix. Transition matrices can be created separately for each listener and block (block matrix), but also for each listener (by pooling data from all blocks of the listener; listener matrix), for a condition (by pooling all data for all listeners for the given condition; condition matrix), and for the whole experiment (by pooling all data of the experiment; experiment matrix). The mean number of switches and the proportions and mean phase durations of the integrated, segregated, and combined percepts were extracted from the transition matrices, separately for each listener and block and also for each listener. Variables for the 'none' response were not analyzed, because the overall proportion of this response was less than 0.8%. The remaining variables characterizing the perception of the tone sequences were used in the analyses to be reported (see below). The transition matrix method enables representing the observed perceptual data in the form of a set of probabilities of transitions from one perceptual state to another and provides a principled way of dealing with the problem of missing observations in a block or listener matrix by substituting missing values from the corresponding condition matrix. As the lack of observing a perceptual state in a given participant/condition does not indicate that participant is not capable of perceiving that pattern, it is important to provide an estimation of the corresponding missing variables in the statistical analysis (e.g., one cannot use the value of 0 for the mean phase duration of an unobserved state, as this distorts the statistical tests).

Data were preprocessed with the help of MATLAB (MathWorks Inc.) scripts. Statistical analyses were conducted with SPSS 15.0 (SPSS Inc.) and R 3.1.2 (R Core Team). All significant effects are reported, together with the $\eta^2$ effect size, observed power and, where applicable, the Greenhouse-Geisser $\varepsilon$ correction factor.

## III. RESULTS

### A. The Number of Blocks Required for Reliably Measuring Idiosyncratic Switching Patterns (Pilot and Main Experiment)

Intra-individual inconsistency was assessed as the Kullback-Leibler divergence (K-L distance) between the block-wise transition matrices of the same listener. Intra-individual inconsistency was then used to determine the number of blocks necessary for detecting individual switching patterns in the pilot experiment. First, K-L distances between all possible pairs of the six experimental blocks of the pilot experiment were calculated to identify the blocks in which the switching pattern substantially differed from that of the rest (see TABLE I for the group-mean within-listener K-L distances between each pair of blocks). Higher values indicate higher intra-individual inconsistency. The pattern of the K-L distances (see TABLE I) suggested that perception in blocks 1 and 4 are the farthest from that observed for the rest of the blocks.

Therefore, mean K-L distances in six possible block-sets were assessed: 1) all six blocks; 2) excluding block 1; 3) excluding block 4; 4) excluding blocks 1 and 4; 5) block 2 and 3, only; and 6) block 5 and 6, only.

Wilcoxon Signed Rank tests were used to determine the block-sets with the best intra-individual consistency. Pairwise comparisons between the all-6 set and the other five sets revealed that only the exclusion of both blocks 1 and 4 improved intra-individual inconsistency compared to that measured from all six blocks ($Z$ = -2.366, $p$ = .018). There was also a strong tendency for improvement when only blocks 5 and 6 were included ($Z$ = -1.859, $p$ = .063). Since blocks 1 and 4 were delivered first within two series of three blocks (see Methods), the perceptual reports recorded for them probably differ from the rest because in these blocks, participants familiarized themselves with the stimuli and the task, after which they could settle into a steady way of performing the task. In other words, blocks 1 and 4 could be regarded as part of the training.

Based on this finding, in the main experiment, we delivered to listeners five blocks without longer breaks in between, and then excluded the responses recorded in the first block from the analysis. This choice was verified *post hoc* in the main experiment similarly as described above. First, the within-listener K-L distances for all block-pairs were averaged across the participants ($M$ = .182, $SD$ = .167). Second, the first block was removed and the remaining within-listener K-L distances averaged ($M$ = .125, $SD$ = .147). A Wilcoxon's Signed Rank test showed that in line with our expectation from the pilot experiment, excluding the first block reduced intra-individual inconsistency ($Z$ = 3.855, $p$ < .001).

**B. Analysis of the Catch Trials (Main Experiment)**

Spearman's rank order correlation was used to determine whether the proportion of time in which the listener's perception matched the percept promoted by the catch trials (catch-trial matching: CTM) was related to their mean intra-individual inconsistency (average of the block-to-block K-L distances of the listener). A significant correlation ($r(52) = -.282$, $p = .043$) was obtained between CTM and the mean intra-individual inconsistency. For illustration purposes, FIG.2 shows a scatterplot of the mean intra-individual inconsistency values ($\log_{10}$ transformed to ensure normal distribution) as a function of the CTM, with the linear regression ($F(1,50) = 4.06$, $p = .049$, $R^2 = 7.52\%$) depicted. A second linear regression was run with excluding participants having a CTM below 60% (see FIG.2). This model was not significant ($F(1,46) = 1.24$, $p = .272$, $R^2 = 2.62\%$). These results shows that, as hypothesized, higher CTM values are associated with lower intra-individual inconsistency – in other words, those participants who showed relatively poor performance on the catch trials (low CTM values) were relatively less consistent in their perceptual reports across the blocks (high intra-individual inconsistency).

For assessing the effects of the participant exclusion criterion, we reassessed the correlation between CTM and the mean intra-individual inconsistency with exclusion based on CTM criteria from 60% to 80% in 5% steps. The number of remaining participants as a function of the exclusion criterion were as follows (correlation between CTM and the mean intra-individual inconsistency and their significance is also given): no exclusion ($N = 52$, $r(52) = -.282$, $p = .043$), 60% exclusion criterion (N = 48, $r(48) = -.171$, $p = .245$), 65% ($N = 34$, $r(45) = -.123$, $p = .420$), 70% ($N = 40$, $r(40) = -.029$, $p = .857$), 75% ($N = 30$, $r(30) = -.214$, $p = .256$), and 80% ($N = 15$, $r(15) = -.100$, $p = .723$). Even employing the lowest (60%) exclusion CTM criterion resulted in the loss of significant correlation between CTM and the mean intra-individual inconsistency (see FIG. 2). This suggests that the correlation was due to a very few (four or fewer) participants, whose inconsistent responses were caused by problems with following the task instructions. In order to preserve the data of as many participants as possible, based on the above results, we employed the 60% cutoff for participant exclusion. The analyses reported in the next sections are thus based on 48 participants' data (77.1% female, $M_{age}$ = 22.5, $SD_{age}$ = 4.2).

**C. Effects of Voluntarily Biasing Switching between Alternative Perceptual Organizations (Main Experiment)**

The effect of voluntary biasing on the average number of switches within the blocks was tested with a Repeated-Measures ANOVA (rmANOVA). Sphericity correction was done by the Greenhouse-Geisser method. The mean number of switches in the Neutral and Hold condition were not normally distributed, thus a $\log_{10}$ transformation was applied in all conditions. The number of switches in the Neutral and the two biased conditions ($M_{neutral}$ = 36.48, $SD_{neutral}$ = 16.44, $M_{hold}$ = 22.49, $SD_{hold}$ = 15.63, and $M_{switch}$ = 49.17, $SD_{switch}$ = 21.36) significantly differed from each other ($F(2,94)=58.654$, $p < .001$, partial $\eta^2 = .555$, observed power > .99, $\varepsilon = .782$). FIG.3 shows that for almost all participants, the number of switches reported with the neutral instructions fell between those reported in the Hold and Switch conditions.

It is possible that participants assumed that each of the perceptual alternatives should be reported with approximately equal probability. This assumption was tested using a Chi-squared test comparing the percentage of the three perceptual alternatives pooled over participants and the four blocks of the neutral condition. Results show that the three values differed from each other ($\chi^2(2) = 10.45$, $p = .005$).

**D. Social Desirability Bias (SDB) Effects (Main Experiment)**

Reliability of the Balanced Inventory of Desirable Responding BIDR scores was measured by Cronbach's alpha. The reliability of the self-deceptive positivity (SDP) factor scores was acceptable (Cronbach's $\alpha$ = .790), whereas it was below the acceptable level for the impression management (IM) factor ($\alpha$ = .562). The magnitude of the SDB effect was estimated by including the two BIDR dimension scores as covariates into the rmANOVA model testing the effect of voluntary biasing. This analysis allows testing the effect of social desirability on the overall outcome of the perceptual experiment. The scale of the BIDR scores and the number of switches differed from each other, thus they were normalized in the following way: the observed minimum value over all participants was subtracted from the individual measures and this difference value was divided by the observed range of the variable (i.e. maximum minus minimum) to preserve the mean differences, separately for of the number of switches (pooled across the three conditions) and the BIDR scores.

There was a significant main effect of task instruction ($F(2,90)=57.981$, $p < .001$, $\eta^2_p$ = .563, observed power > .99, $\varepsilon$ = .763). Neither SDP ($F(2,90)=1.007$, $p$ = .369, $\eta^2_p$ = .022, observed power = .196) nor IM ($F(2,90)=0.106$, $p$ = .847, $\eta^2_p$ = .002, observed power = .064) showed significant interactions with the main effect of task instructions.

In addition, the effects of SDB were tested with independent correlations between the BIDR dimensions and the auditory perceptual variables extracted from the listener transition matrices as well as intra-individual inconsistency obtained for the neutral instruction set for testing possible effects of the social desirability bias on variables extracted from the subjective perceptual reports in the main (neutral) condition. Spearman's Rank Order correlations were used, because intra-individual inconsistency, the three mean duration variables, and the proportion of segregated were not normally distributed. This approach is deliberately lenient: as the tests are not independent of each other, correction by family-wise error would offer a more conservative assessment. However, we wished to detect any possible effect of SDB in order to uncover any possible source of bias in the subjective reports.

None of the correlations between any of the perceptual variables or the intra-individual inconsistency and the two SDB dimensions reached significance (see TABLE II.) and only the correlations between SDP and the proportion of the segregated responses showed a strong tendency, suggesting that the participants with higher SDP tended to report hearing the segregated percept for a smaller proportion of time than the ones with lower SDP.

**E. Differences between Random Button Presses and Perceptual Reports (Follow-up experiment)**

Distributions of the button press durations (pooling across the different button combinations) were created separately for the Random/Auditory × Neutral/Hold/Switch conditions (FIG. 4). First, all distributions were tested against the gamma and lognormal distributions using One-Sample Kolmogorov-Smirnov tests (Massey, 1951). Results (TABLE III) showed that only the Auditory-Hold condition was consistent with a gamma distribution, whereas the other two conditions produced significantly different distributions. On the other hand, all auditory conditions could be described with the lognormal distribution. In contrast, none of the random conditions could be described by either of these distributions. Further, employing Two-Sample Kolmogorov-Smirnov tests we found that the distributions observed in the random and auditory conditions differed from each other under each of the neutral/voluntary-bias instructions (TABLE III). Finally, an rmANOVA with the factors of (Random/Auditory × Neutral/Switch/Hold) was conducted on the number of switches. Sphericity violation correction was done using the Greenhouse-Geisser method. Because the variables were not normally distributed, $\log_{10}$ transformation was applied to them. Significant main effects were observed for the Random/Auditory ($F(1,13)=7.137$, $p = .019$, $\eta^2_p = .354$, observed power = .695, $\varepsilon = 1$) and the Neutral/Switch/Hold factors ($F(2,26)=47.289$, $p < .001$, $\eta^2_p = .785$, observed power > .99, $\varepsilon = .767$), whereas the interaction between the two factors was not significant ($F(2,26)=3.182$, $p = .078$, $\eta^2_p = .197$, observed power = .464, $\varepsilon = .719$). Considerably higher number of switches occurred in the random than in the auditory conditions and the biasing instructions had the expected Hold<Neutral<Switch effect in both cases (TABLE III).

## IV. DISCUSSION

The aim of the pilot experiment was to test whether it is possible to record a set of subjective reports with a relatively simple experimental procedure (compared to Denham et al., 2014) with high consistency within individuals. Data from the pilot experiment suggested that adequate intra-individual consistency can be achieved by recording five four-minute-long stimulus blocks successively within a single session, with the first stimulus block being regarded as a training block for which the reports are not analyzed. The main experiment confirmed that the exclusion of the first block from the analysis increases the intra-individual consistency across blocks. This procedure allows studying the consistency and inter-individual variability of perception in a multi-stable auditory streaming paradigm, as was demonstrated by Denham et al. (2014), without the need to employ a time-consuming multi-session experimental design.

The follow-up experiment confirmed that listeners produce markedly different reports when instructed to mark their perception of the sounds than when instructed to press buttons randomly. Further, similarly to Pressnitzer and Hupé's (2006) results, we found that the distribution of button presses produced when listeners are instructed to mark their perceptions in response to a bi- or multi-stable tone sequence can be best approximated by the lognormal distribution, suggesting that these responses are produced by independent stochastic processes (Murata et al., 2003). These results suggest that listeners do not disregard the instructions to mark their perceptions, even though the experimenter has no way of directly verifying the truthfulness of their responses.

In the following paragraphs, the results related to the three potential biasing factors are discussed. When recording subjective reports, it is especially important to find a way to exclude participants who do not reliably follow the instructions. There are several possible reasons for such behavior: the participant may not have fully understood or remembered the instructions; fatigue and not paying attention to the task could also lead to unreliable data. Catch trials, for which the expected report is known, can help to detect such behavior. We found that higher catch trial matching (CTM) values were associated with higher intra-individual consistency. One possible reason for this was that participants with lower CTM reported their perception in an inconsistent manner. To test this possibility, we reassessed the correlation between CTM and intra-individual inconsistency by excluding participants with low CTM. We found that even excluding only a few participants this way (those with a CTM below 60%) eliminated the significant correlation between CTM and intra-individual inconsistency. This suggests that the significant correlation was to a large degree caused by a few participants with low CTM and low intra-individual consistency. That is, participants with higher CTM values can be regarded as a reasonably homogenous and reliable group.

Results obtained with the voluntary biasing instructions showed that, in general, participants were able to bias their perception such that it switched between the alternatives more (or less) often than under the neutral instructions. The follow-up experiment provided convergent results. The results rule out a number of possible voluntary biasing scenarios for the subjective reports obtained with the neutral instructions: listeners 1) did not randomly press the response buttons; 2) did not attempt to switch as often or as seldom as they could; and 3) did not follow a strategy to mark the three possible response types in approximately equal proportion. Although there is no way to rule out that some or possibly even all listeners biased their perception by idiosyncratic voluntary strategies, it appears likely that the experimental procedures did not promote a systematic bias of the subjective perceptual reports. The results obtained in the biased conditions are also consistent with those previously reported for bistable auditory (Pressnitzer & Hupé, 2006) and visual stimulus configurations (e.g., Klink et al., 2008; Toppino, 2003; van Ee et al., 2005). This result thus demonstrates a further similarity between the perception of auditory and visual bistable stimuli (see Denham et al., 2013; Pressnitzer & Hupé, 2006; Schwartz et al., 2012).

Sensitivity of the subjective reports in the auditory streaming paradigm to potential biases caused by perceived (implicit) expectations present in the experimental situation was tested by measuring participants' social desirability bias (SDB, Paulhus, 1984, 1991). The assumption underlying this approach was that biases in the subjective reports caused by perceived implicit expectations would be stronger for participants who are more susceptible to SDB. The results showed that neither of the social desirability dimensions (self-deceptive positivity and impression management) was associated with any of the perceptual variables or with the outcome of the voluntary bias manipulations. The null results were obtained despite the rather liberal statistical approach (i.e., correlation tests were treated as being independent of each other without family-wise error or similar corrections being employed). Given that SDB measures sensitivity to expectations including implicit ones, the lack of significant association between SDB and the perceptual variables suggests that participants did not significantly misreport their perceptions due to perceived implicit expectations. In sum, these results support the validity of subjective reports within the current experiment.

If the level of susceptibility to SDB was low in our sample (compared to the overall population), it would explain the lack of observed SDB effects. TABLE IV. shows side by side the mean SDB scores obtained from a sample of Hungarian university students ($N$ = 708, 72.3% female, age range between 18 and 34 years, $M_{age}$ = 22.94, $SD_{age}$ = 3.28; personal communication by Gábor Orosz) with those obtained in the current experiment. The two studies used identical SDB measurements. For both SDB dimensions, the range, mean, and standard deviation are quite similar for the two groups. Although a direct comparison between the two samples is not possible due to the large difference in the sample size, it is reasonable to suggest that the current sample did not substantially differ from the larger one. Thus the explanation of the lack of SDB effect being due to the small sample size is unlikely.

Self-deceptive positivity is defined as a general biasing tendency working within an individual, whereas impression management is dependent on the context (Paulhus, 1991; Zerbe & Paulhus, 1987). Taking this into account, the lack of significant SDB effects suggests that neither the experimental context nor the task itself generated strong and systematic implicit expectations that caused participants to alter their natural behavior by a detectable amount. The instructions emphasized that there were no ‘correct’ or ‘incorrect’ responses in the task as the aim of the study was to learn how each individual experiences these stimuli. This may have created a friendly, non-stressful experimental atmosphere. The main limitation for this interpretation is that impression management had low reliability within our sample. Future studies could further investigate the effects of task instructions by collecting data from a group of participants to whom some implicit expectations have been communicated.

In sum, the current study constitutes a reasonably simple paradigm for studying intra-individual consistency and inter-individual differences in auditory multistable perception. Perceptual reports differed from random button presses, thus it seems likely that listeners responded to the experimental stimuli. Further, we have tested three major sources that could possibly bias subjective perceptual reports collected within the auditory streaming paradigm. We found that the proportion of time during which participants’ reports matched the perception promoted by catch trial sequences successfully detected those participants who did not reliably follow the task instructions. Rejecting them made the data from the remaining group sufficiently reliable. We also showed that participants are not likely to have biased their perception or their report of it either voluntarily or in response to perceived implicit expectations of the experimental situation. These results show that by careful planning of the stimuli and the task as well as the experimental situation (such as the wording of task instructions) it is possible to obtain valid data using subjective reports. Because for certain research questions subjective reports have important advantages over so-called ‘objective’ measures, perceptual research benefits from the possibility of recording reliable and valid subjective report data.

## ACKNOWLEDGEMENTS

This study was supported by the Hungarian Academy of Sciences (Lendület Project LP-36/2012). The experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience. The authors thank one of the anonymous reviewers for suggesting the important follow-up experiment.

**FOOTNOTES**

[1] This test was suggested by one of the reviewers.

[2] It is important to note that social desirability is also measured by a self-report scale. Thus the responses can, in principle, be faked. However, it is unlikely that a substantial number of the listeners in our experiment had the desire and the necessary knowledge to systematically fake the responses to this questionnaire.

[3] The following further tasks and questionnaires were completed by the participants. First, they completed a test battery containing the Ego-resiliency scale (Block & Kremen, 1996), the Big Five Inventory (John & Srivastava, 1999), the UPPS Impulsive Behavior Scale (Whiteside & Lynam, 2001), the NISROE, which measures internal and external encoding styles (Lewiczki, 2005), and finally, the BIDR questionnaire (see main text). They then performed some tasks testing executive functions. These included a computerized version of the Stroop task (Stroop, 1935), a computerized N-back task (Kirchner, 1958), and a semantic fluency task (Troyer, Moscovitch, & Winocur, 1997). Finally, participants performed tasks assessing creativity, which consisted of the Use of Objects Task and a caption generation task (Jung, Grazioplene, Caprihan, Chavez, & Haier, 2010). The related results will be published elsewhere.

## REFERENCES


Bendixen, A. **(2014).** „Predictability effects in auditory scene analysis: A review", Front. Neurosci. **8**(60). doi:10.3389/fnins.2014.00060

Bendixen, A., Denham, S. L., Gyimesi, K., and Winkler, I. **(2010).** „Regular patterns stabilize auditory streams", J. Acoust. Soc. Am., **128**(6), 3658–3666. doi:10.1121/1.3500695.

Block, J., and Kremen, A. M. **(1996).** „IQ and Ego-resiliency: Conceptual and empirical connections and separateness", J. Pers. Soc. Psychol. **70**(2), 349–361. doi:10.1037/0022-3514.70.2.349

Bregman, A. S. **(1990).** *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, Cambridge, MA), pp. 792.

Deike, S., Heil, P., Böckmann-Barthel, M., and Brechmann, A. **(2012).** „The build-up of auditory stream segregation: A different perspective", Front. Psychol. **3**(461). doi:10.3389/fpsyg.2012.00461

Denham, S. L., Bendixen, A., Mill R., Tóth, D., Wennekers, T., Coath, M., Bőhm, T. T., Szalárdy, O.; and Winkler, I. **(2012).** „Characterising switching behaviour in perceptual multi-stability", J. Neurosci. Methods. **210**(1), 79–92. doi:10.1016/j.jneumeth.2012.04.004.

Denham, S. L., Bőhm, T. M., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill R., and Winkler, I. **(2014).** „Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli", Front. Neurosci. **8**(25). doi:10.3389/fnins.2014.00025.

Denham, S. L., Gyimesi, K., Stefanics, I., and Winkler, I. **(2010).** „Stability of perceptual organisation in auditory streaming", *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Plamer, and R. Meddis (Springer, New York, NY), pp. 477–488.

Denham, S. L., Gyimesi, K., Stefanics, I., and Winkler, I. **(2013).** **”**Perceptual bistability in auditory streaming: How much do stimulus features matter?”, Learning and Perception. **5**(S2), 73–100.

Denham, S. L., and Winkler, I. **(2006).** „The role of predictive models in the formation of auditory streams”, J. Physiol. Paris. **100**(1-3), 154–170. doi:10.1016/j.jphysparis.2006.09.012

Fisher, R. J. **(1993).** „Social desirability bias and the validity of indirect questioning”, J. Consum. Res. **20**(2), 303–315. doi:10.1086/209351

Fisher, R. J., and Katz, J. E. **(2000). „**Social-desirability bias and the validity of self-reported values”, Psychol. Market. **17**(2), 105–120. doi:10.1086/209351

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. **(2005).** „Neuromagnetic correlates of streaming in human auditory cortex”, J. Neurosci. **25**(22), 5382–5388. doi:10.1523/JNEUROSCI.0347-05.2005

John, O., and Srivastava, S. **(1999).** „The Big Five trait taxonomy: History, measurement, and theoretical perspectives”. *Handbook of Personality: Theory and Research,* edited by L. A. Pervin and O. John (The Guilford Press, New York, NY), pp. 102–138.

Jung, R. E., Grazioplene, R., Caprihan, A., Chavez, R. S., and Haier, R. J. **(2010).** „White matter integrity, creativity, and psychopathology: Disentagling constructs with diffusion tensor imaging”, PLoS ONE. **5**(3), e9818. doi:10.1371/journal.pone.0009818.

Kashino, M., Kondo, H. M., and Okada, M. **(2008). „**Perceptual dynamics of auditory streaming and its neural correlates”, J. Acoust. Soc. Am. **123,** 3052. doi:10.1121/1.2932766

King, M. F., and Bruner, G. C. **(2000).** „Social desirability bias: A neglected aspect of validity testing”, Psychol. Market. **17**(2), 79–103. doi:10.1002/(SICI)1520-6793

Kirchner, W. K. **(1958).** „Age differences in short-term retention of rapidly changing information", J. Exp. Psychol. **55**(4), 352–358. doi:10.1037/h0043688

Klink, P. C., van Ee, R., Nijs, M. M., Brouwer, G. J., Noest, A. J., and van Wezel, R. **(2008).** „Early interactions between neuronal adaptation and voluntary control determine perceptual choices in bistable vision", J. Vis. **8**(16), 1–18. doi:10.1167/8.5.16.

Leopold, D.A. **(1997).** "Brain mechanisms of visual awareness", PhD thesis, Baylor College of Medicine, Houston, TX.

Lewiczki, P. **(2005).** „Internal and external encoding stlye and social motivation". In J. P. Forgas, *Social Motivation: Conscious and Unconscious Processes,* edited by K. D. Williams and S. M. Laham (Cambridge University Press, New York, NY), pp. 194–209.

Massey, F. J. **(1951).** "The Kolmogorov-Smirnov Test for Goodness of Fit." J. Am. Stat. Assoc. **46**(253), 68–78. doi: 10.2307/2280095

Moore, B. C., and Gockel, H. E. **(2012).** „Properties of auditory stream formation", Philos. Trans. R. Soc. Lond. B. Biol. Sci. **367**(1591), 919–931. doi:10.1098/rstb.2011.0355.

Moreno-Bote, R., Shpiro, A., Rinzel, J., and Rubin, N. **(2010).** „Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance", J. Vis. **10**(11), 1. doi:10.1167/10.11.1.

Murata, T., Matsui, N., Miyauchi, S., Kakita, Y., & Yanagida, T. **(2003).** "Discrete stochastic process underlying perceptual rivalry", Neuroreport. **14**(10), 1347-1352. doi: 10.1097/00001756-200307180-00013

Paulhus, D. L. **(1984).** „Two-component models of socially desirable responding". J. Pers. Soc. Psychol. **46**(3), 598–609. doi:10.1037/0022-3514.46.3.598

Paulhus, D. L. **(1991).** „Measurement and control of response bias", *Measures of Personality and Social Psychological Attitudes*, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman (Academic Press, San Diego, CA), pp. 17–59.

Pressnitzer, D., and Hupé, J.-M. **(2006).** „Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization", Curr. Biol. **16**(13), 1351–1357. doi:10.1016/j.cub.2006.05.054

Schwartz, J. L., Grimault, N., Hupé, J. M., Moore, B. C., and Pressnitzer, D. **(2012).** **„**Multistability in perception: Binding sensory modalities, an overview", Philos. Trans. R. Soc. Lond. B. Biol. Sci. **367**(1591), 896–905. doi:10.1098/rstb.2011.0254.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., and Alain, C. **(2012).** „Attention, awareness, and the perception of auditory scenes", Front. Psychol. **3**(15). doi:10.3389/fpsyg.2012.00015

Stöber, J., Dette, D. E., and Musch, J. **(2002).** „Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding", J. Pers. Assess. **78**(2), 370–389. doi:10.1207/S15327752JPA7802_10

Stroop, J. R. **(1935).** „Studies of inference in serial verbal reactions", J. Exp. Psychol. **18**(6), 643–662. doi:10.1037/h0054651

Toppino, T. **(2003).** „Reversible-figure perception: Mechanisms of intentional control", Percept. Psychophys. **65**(8), 1285–1295. doi:10.3758/BF03194852

Troyer, A. K., Moscovitch, M., and Winocur, G. **(1997).** „Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults", Neuropsychology. **11**(1), 138–146. doi:10.1037//0894-4105.11.1.138

van Ee, R., van Dam, L., and Brouwer, G. J. **(2005).** „Voluntary control and the dynamics of perceptual bi-stability", Vision. Res. **45**(1), 41–55. doi:10.1016/j.visres.2004.07.030

van Noorden, L. **(1975).** *Temporal coherence in the perception of tone sequences* (Technical University of Eindhoven, Eindhoven), Doctoral dissertation, pp. 127.

Whiteside, S. P., and Lynam, D. R. **(2001).** „The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity", Pers. Indiv. Diff. **30**(4), 669–689. doi:10.1016/S0191-8869(00)00064-7

Zerbe, W. J., and Paulhus, D. L. **(1987).** „Socially desirable responding in organizations: A reconception", Acad. Manage. Rev. **12**(2), 250–264. doi:10.5465/AMR.1987.4307820

Zhou, Y. H., Gao, J. B., White, K. D., Merk, I., & Yao, K. **(2004).** "Perceptual dominance time distributions in multistable visual perception", Biol. Cybern. **90**(4), 256-263. doi: 10.1007/s00422-004-0472-8

TABLE I. Group-mean (n=7) within-participant Kullback-Leibler distances with standard deviations in parentheses between the transition matrices computed for each of the six blocks in the pilot experiment

| Blocks | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 2 | .15 (.12) | | | | |
| 3 | .33 (.49) | .12 (.14) | | | |
| 4 | .21 (.14) | .11 (.11) | .09 (.08) | | |
| 5 | .39 (.43) | .14 (.11) | .08 (.09) | .18 (.27) | |
| 6 | .35 (.48) | .12 (.12) | .07 (.08) | .19 (.26) | .06 (.05) |

TABLE II. Spearman's rank correlations between perceptual variables and the two dimensions of social desirability in the main experiment

| | SDP | | IM | |
|---|---|---|---|---|
| | rho | *p* | rho | *p* |
| Duration of integrated | -.059 | .688 | -.048 | .746 |
| Duration of segregated | -.161 | .273 | -.010 | .945 |
| Duration of combined | .074 | .619 | .051 | .728 |
| Proportion of integrated | .003 | .983 | -.108 | .465 |
| Proportion of segregated | -.273 | .061 | -.063 | .762 |
| Proportion of Combined | .239 | .102 | .071 | .631 |
| Mean number of switches | .147 | .320 | .027 | .853 |
| Intra-individual inconsistency | .162 | .270 | .174 | .236 |

*notes:* rho = Spearman's rank correlation coefficient, *p* = p-value of the correlation, SDP = self-deceptive positivity, IM = impression management

TABLE III. Group-average (n=14) button-press durations, the results of comparisons against the gamma and lognormal distributions, and between the Auditory and Random conditions in the follow-up experiment. The standard deviation is included next to the mean value in parenthesis. Comparison to the gamma and lognormal distributions were done with one-sample Kolmogorov-Smirnov tests: the *p*-value is included in a parenthesis next to the *D*-value. The last row provides the results of the two-sample Kolmogorov-Smirnov tests between the Auditory and the Random condition with the *p*-value included in a parenthesis next to the *D*-value.

| | | Neutral | Switch | Hold |
|---|---|---|---|---|
| Auditory | mean(SD) | 30.76 (26.22) | 52.05 (29.64) | 17.24 (11.73) |
| | gamma | .096 (.001) | .062 (.008) | .078 (.136) |
| | lognormal | .055 (.175) | .034 (.367) | .071 (.221) |
| Random | mean(SD) | 95.84 (92.11) | 181.55 (185.54) | 52.46 (91.92) |
| | gamma | .167 (<.001) | .224 (<.001) | .215 (<.001) |
| | lognormal | .140 (<.001) | .160 (<.001) | .164 (<.001) |
| Comparison | | .466 (<.001) | .603 (<.001) | .445 (<.001) |

TABLE IV. Descriptive statistics of the two social desirability dimensions in the main experiment and a large Hungarian university student sample for reference

| | Self-deceptive positivity | | Impression management | |
|---|---|---|---|---|
| | Current sample ($N$ = 48) | University sample ($N$ = 708) | Current sample ($N$ = 48) | University sample ($N$ = 708) |
| Mean | 81.35 | 86.42 | 86.58 | 78.48 |
| Standard deviation | 13.30 | 9.86 | 11.95 | 8.74 |
| Minimum | 55 | 56 | 67 | 48 |
| Maximum | 116 | 125 | 111 | 106 |

FIG.1. Schematic depiction of the auditory streaming paradigm (top panel) and its possible perceptions grouped into 3 categories (as was done in the current study and by e.g. Bendixen et al., 2010; the 3 lower panels). Rectangles depict the ‘A’ and ‘B’ sounds with feature difference between them indicated by displacement in the vertical direction. Time flows along the horizontal direction. Sounds perceived as part of the same stream are connected by lines. Darker notes with grey background indicate the stream in the foreground (also described with symbols to the right of each lower panel).

FIG.2. Linear regression of the $\log_{10}$ mean intra-subject inconsistency (y axis) over catch trial matching (CTM; x axis) in the main experiment. Solid line represents the regression over all participants, whereas dashed line that after employing the 60% exclusion criterion and thus rejecting the data of the four listeners with the lowest CTM values (marked with full circles).

FIG.3. Average number of switches in the Neutral (circles), Hold (squares), and Switch (triangles) conditions, separately for each participant (1 to 48, empty symbols) and for the whole group (n=48; filled symbols; right side) in the main experiment.

FIG.4. Density distributions of button press durations in the six conditions (Auditory/Random × Neutral/Switch/Hold) in the follow-up experiment. Please note that the top of the peak for the Random/Switch condition is not visible, because adjusting the y-axis to it’s high value (1.2) would reduce the resolution of the traces on the other panels.

FIG.1.

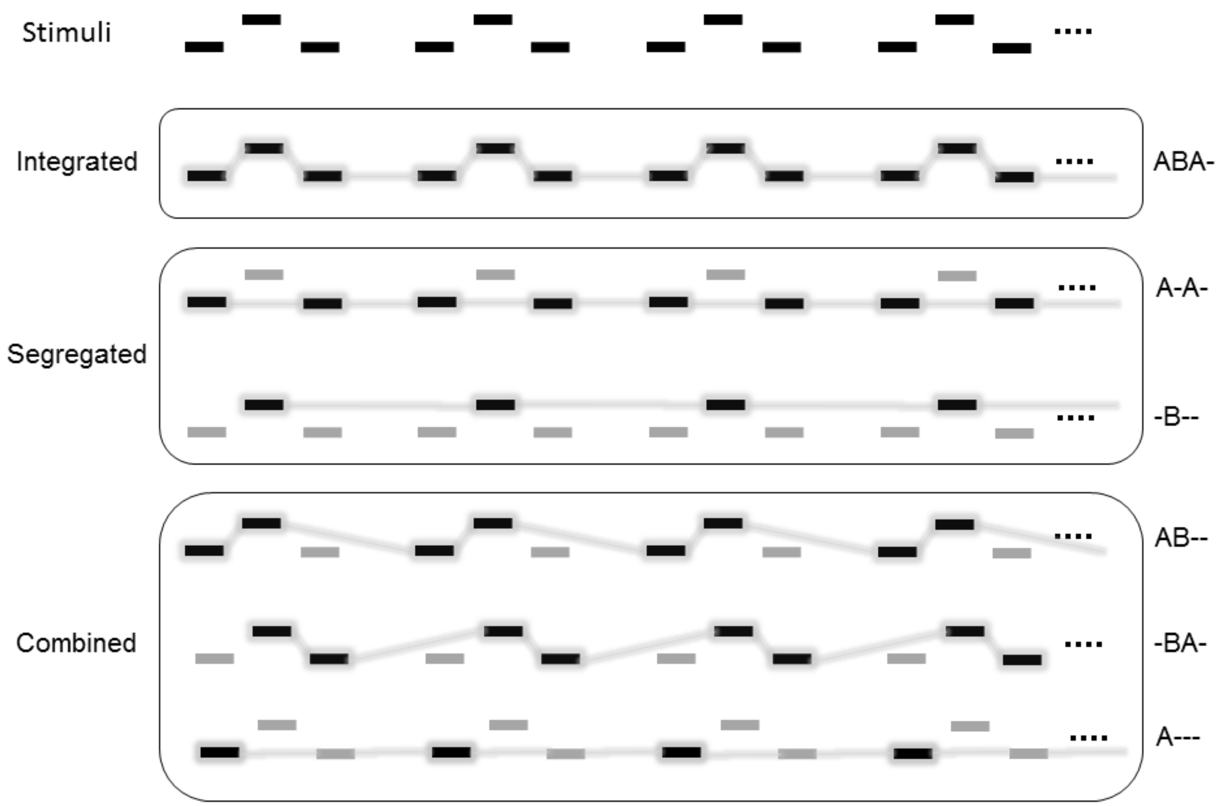
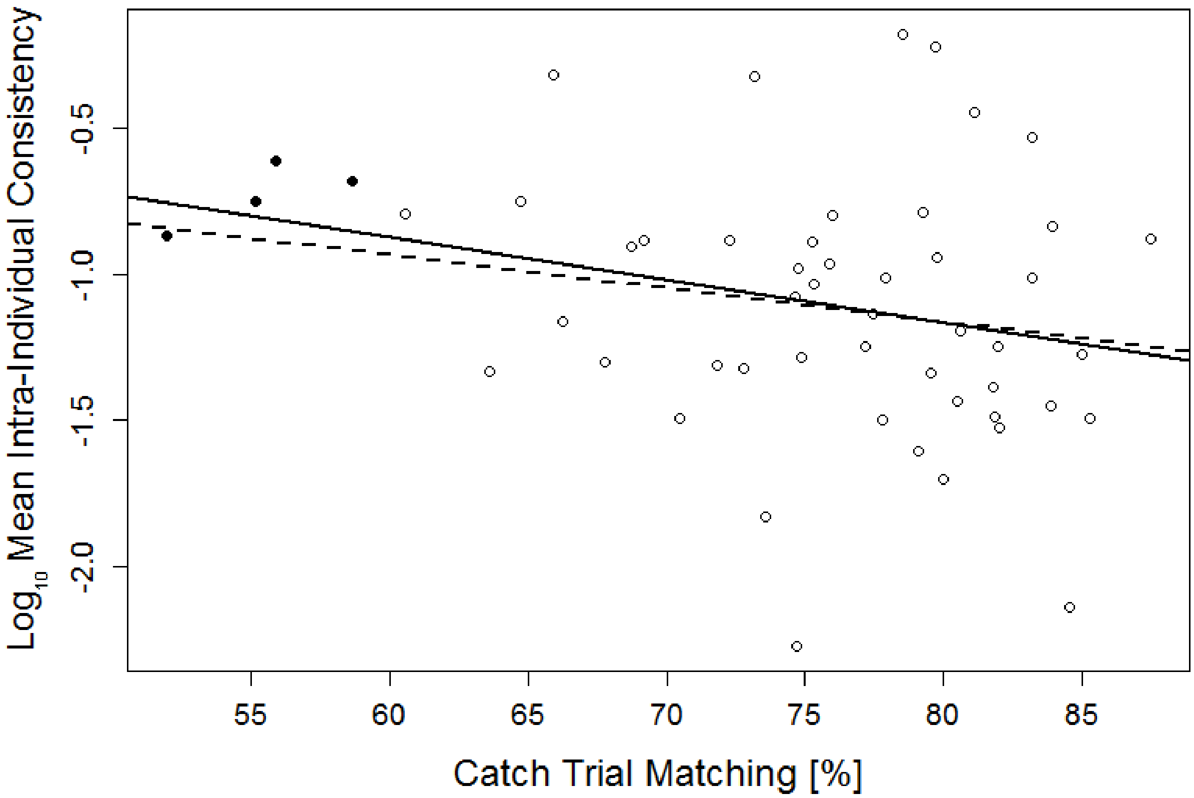


FIG.2



Log10 Mean Intra-Individual Consistency
-0.5
-1.0
-1.5
-2.0
55
60
65
70
75
80
85
Catch Trial Matching [%]

FIG.3

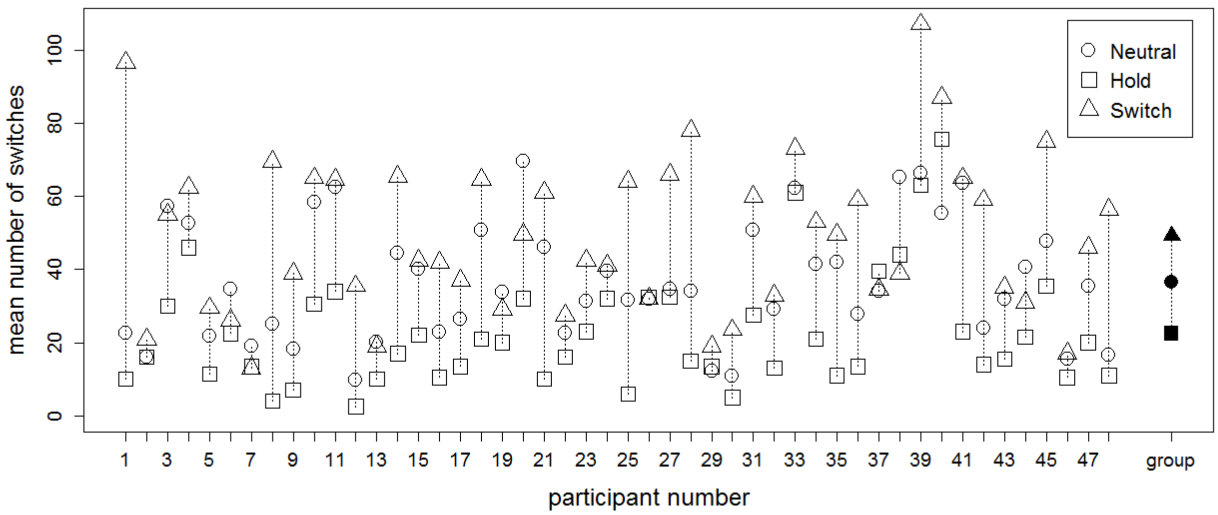
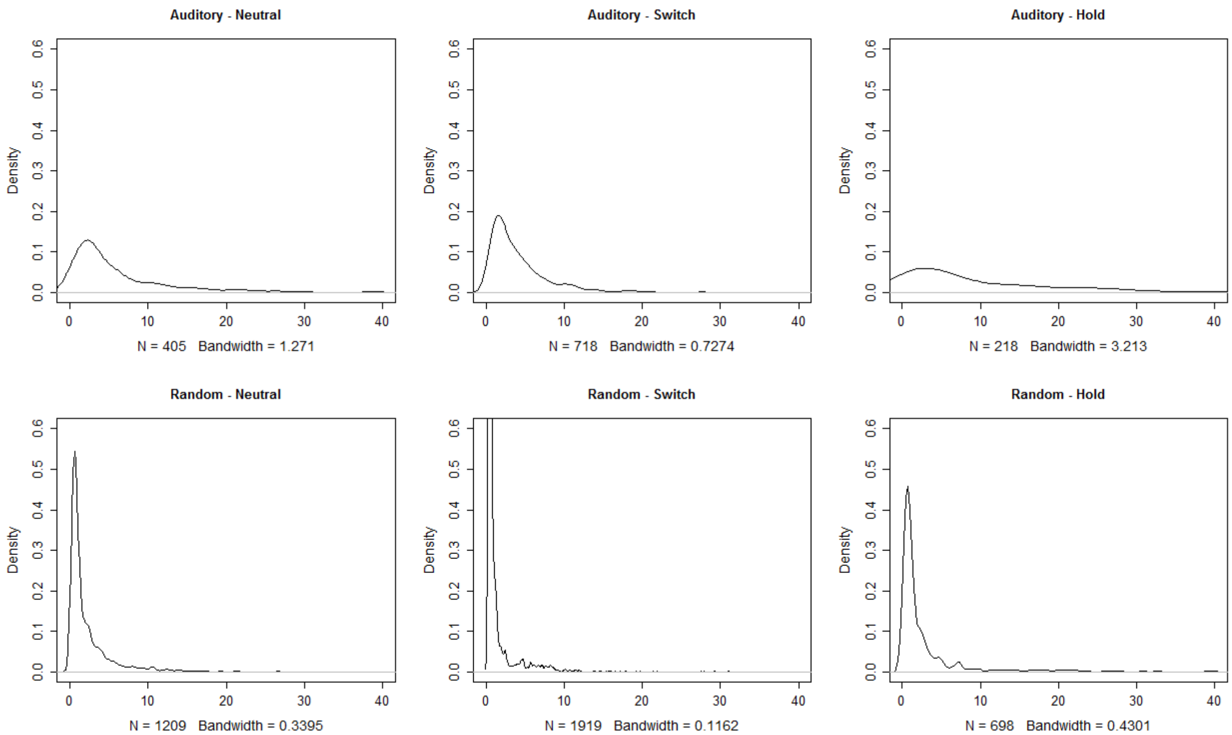

mean number of switches
0
20
40
60
80
100
participant number
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
35
37
39
41
43
45
47
group
Neutral
Hold
Switch

FIG.4


Auditory - Neutral
N = 405 Bandwidth = 1.271
Auditory - Switch
N = 718 Bandwidth = 0.7274
Auditory - Hold
N = 218 Bandwidth = 3.213
Random - Neutral
N = 1209 Bandwidth = 0.3395
Random - Switch
N = 1919 Bandwidth = 0.1162
Random - Hold
N = 698 Bandwidth = 0.4301
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
10
20
30
40