# Change Detection in Auditory Textures

**Yves Boubenec, Jennifer Lawlor, Shihab Shamma and Bernhard Englitz**

**Abstract** Many natural sounds have spectrotemporal signatures only on a statistical level, e.g. wind, fire or rain. While their local structure is highly variable, the spectrotemporal statistics of these auditory textures can be used for recognition. This suggests the existence of a neural representation of these statistics. To explore their encoding, we investigated the detectability of changes in the spectral statistics in relation to the properties of the change.

To achieve precise parameter control, we designed a minimal sound texture—a modified cloud of tones—which retains the central property of auditory textures: solely statistical predictability. Listeners had to rapidly detect a change in the frequency marginal probability of the tone cloud occurring at a random time.

The size of change as well as the time available to sample the original statistics were found to correlate positively with performance and negatively with reaction time, suggesting the accumulation of noisy evidence. In summary we quantified dynamic aspects of change detection in statistically defined contexts, and found evidence of integration of statistical information.

**Keywords** Auditory textures · Change detection · Sound statistics

Y. Boubenec (✉) · J. Lawlor · S. Shamma · B. Englitz
Laboratoire des Systèmes Perceptifs, CNRS UMR 8248, 29 rue d'Ulm, 75005 Paris, France
e-mail: boubenec@ens.fr

Y. Boubenec · J. Lawlor · S. Shamma · B. Englitz
Département d'études cognitives, Ecole normale supérieure PSL Research University,
29 rue d'Ulm, 75005 Paris, France

S. Shamma
Neural Systems Laboratory, University of Maryland in College Park, MD, USA

B. Englitz
Department of Neurophysiology, Donders Centre for Neuroscience, Radboud Universiteit
Nijmegen, Nijmegen, The Netherlands



P. van Dijk et al. (eds.), *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing,* Advances in Experimental Medicine and Biology 894,
DOI 10.1007/978-3-319-25474-6_24

# 1 Introduction

For many natural sounds, such as wind, fire, rain, water or bubbling, integration of statistical information under continuous conditions is critical to assess the predictability of changes. While their spectrogram is dominated by small, recognizable elements (e.g. gust, crackles, drops), their occurrence is not predictable *exactly*, but they occur randomly in time and frequency, yet constrained by certain probabilities. As shown previously, these probabilities/statistics define the identity of such sounds, so called auditory textures (McDermott et al. 2013). McDermott and others could demonstrate this point directly, by recreating auditory textures from samples, using only the spectrotemporal statistics of those sounds (e.g. McDermott and Simoncelli 2011).

Detecting relevant changes in auditory textures is hence complicated due to the lack of deterministic spectrotemporal predictability, and instead relies on the prediction of stimulus statistics (and their dynamics). The aim of this study is to determine how changes in stimulus statistics are detected under continuous presentation of stimulus. Concretely, we assessed how first-order non-uniform sound statistics are integrated and how listeners disentangle a change in these from the intrinsic stochastic variations of the stimulus.

We designed a broadband stimulus, composed of tone-pips, whose occurrence is only constrained by a frequency marginal probability. While the central 'textural' property of solely statistical predictability is maintained, it is devoid of most other systematic properties, such as frequency modulation, across channel or temporal correlations. A change is introduced by modifying the marginal distribution at a random time during the stimulus presentation, which subjects are asked to report.

This change detection task allows us to explicitly address the integration of statistical information under continuous conditions, which models more closely the real-world challenge of detecting changes in complex ongoing auditory scenes. In this context, we studied the dependence of detection on the time and size of the change. We found evidence for integration of statistical information to predict a change in stimulus statistics, both based on hit rate as well as on reaction times. The main predictors appear to be the time available for integration of evidence.

# 2 Methods

## *2.1 Participants*

Fifteen subjects (mean age 24.8, 8 females, higher education: undergraduate and above) participated in the main study in return for monetary compensation. All subjects reported normal hearing, and no history of psychiatric disorders.

## 2.2 Experimental Setup

Subjects were seated in front of a screen with access to a response box in sound-attenuated booth (Industrial Acoustics Company GmbH). Acoustic stimulus presentation and behavioural control were performed using custom software package written in MATLAB (BAPHY, NSL, University of Maryland). The acoustic stimulus was sampled at 100 kHz, and converted to an analog signal using an IO board (National Instruments NI PCIe-6353) before being sent to diotic presentation using headphones (Sennheiser HD380, calibrated flat, i.e. ±5 dB within 100 Hz–20,000 Hz). Reaction times were measured via a custom-built response box and collected by the same IO card with a sampling rate of 1 kHz.

## 2.3 Stimulus Design

We used a simplified sound texture model, which retained the property of being predictable only from the statistical properties of its complex spectrotemporal structure (Fig. 1a). The texture was a tone cloud consisting of a sequence of 30-ms, temporally overlapping pure tones whose frequency covered a range of 2.2 octaves (400–1840 Hz) divided in 8 frequency bins. The frequency resolution of the tone distribution was 12 semitones per octave, starting at 400 Hz. This allowed 26 tone frequencies in total with 3–4 frequency values in each frequency bin.

The minimal temporal unit of the stimulus was a 30-ms chord in which the number of tones in a particular frequency range was drawn according to a marginal distribution for that frequency range. On average, for each chord duration the mean number of tones per octave was 2. The marginal distribution of occurrence probability was obtained by modifying a uniform distribution in each frequency bin. The probability in each of these is called $P_{uniform} = 1/8 = 0.125$. To generate different initial frequency marginals, we perturbed the marginals randomly by adding/subtracting fixed value $\Delta$ corresponding to 50% of the probability (which would be observed for a uniform distribution ($\Delta = P_{uniform}/2 = 0.0625$)). The resulting marginal distribution was thus pseudo-random with 3 bins at $P_{uniform}+\Delta$, 3 bins at $P_{uniform}-\Delta$ and 2 bins left intact at $P_{uniform}$. This implies that the average initial probability in 2 bins can take 5 different values, namely $\Delta$, $3\Delta/2$, $2\Delta$, $5\Delta/2$, and $3\Delta$.

The change consisted in an increment of the marginal distribution in the selected frequency bins at a random point in time (referred to as *change time*) during stimulus presentation. We chose to use an increment of the marginal distribution in two adjacent frequency bins on the basis that an appearance is more salient than its opposite in a complex acoustic stimulus (Constantino et al. 2012). After the change, the second stimulus continued for up to 2 s or until the subject made an earlier decision, whichever happened first.

The increment size, referred to as *change size*, was drawn from a set of discrete values [50, 80, 110, 140] %, relative to the single bin probability in a uniform

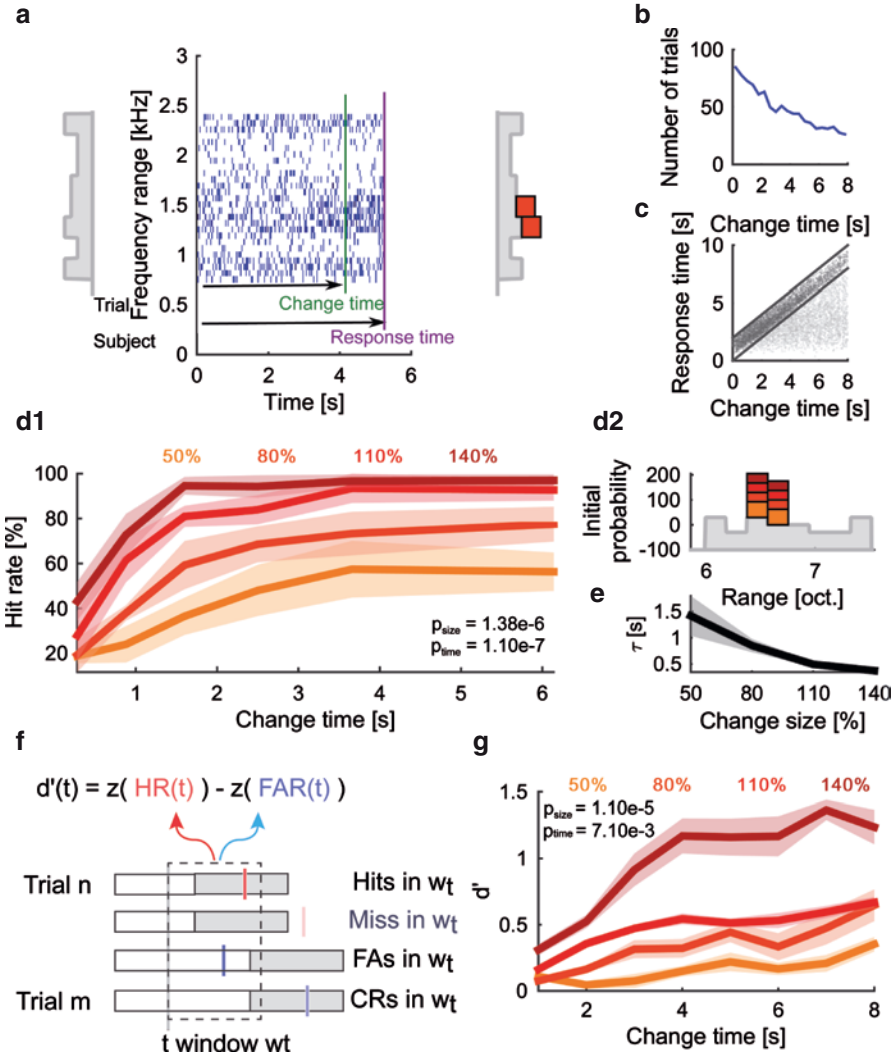


**Fig. 1** Detection of change in statistics requires integration over time. **a** Subjects listened to an acoustic textural stimulus, whose sole predictability was governed by its marginal frequency distribution (*grey curve, left*). Tones in individual frequency bins were all drawn independent consistent with the marginal (*middle*). Listeners were instructed to report changes by a button press. The frequency marginal was modified after a randomly chosen point in time ('change time', *red* bars on the *right*). The probability in two frequency bins separated by a random distance was modified at a time. All other bins were slightly reduced in probability to maintain the overall loudness constant. **b** Distribution of change times across all subjects ($n = 10$). Change times were drawn from an exponential distribution. **c** Response times occurred before (false alarms) and after the change time (hits). Subjects usually responded only after an initial time of listening, allowing them to acquire the acoustic statistics. **d** Hit rate of change detection depended significantly on change time (x-axis) and change size (colour). **e** Four different change sizes were used (colour-coded). **f** The dynamics of the hit rate curve also varied with change size, indicated by the fitted parameter $\tau$ of a cumulative Erlang distribution (see *Methods*). **g** To control for false alarms, d' was also computed as a function of time into the trial (see *Methods* for details), where within each trial the statistics change where the gray area starts. **h** Similar to hit rate, d' increased significantly over time into the trial, and with change size. Errorbars indicate confidence bounds two SEM based on a bootstrap across subjects

distribution (for 8 bins, the single bin uniform probability is 1/8th, and thus a 50% change size would be 1/16th). In order to maintain the overall number of tones per chord before and after the change, the occurrence probability in the six remaining frequency bins was decreased accordingly.

The time at which the change occurred (*change time*) was drawn randomly from an exponential distribution (mean: 3.2 s) limited to the interval [0, 8] s. This choice of distribution prevents subjects from developing a timing strategy, as the instantaneous probability of a change in the next time step is constant.

## 2.4 Procedure

The experiment was separated into three phases: instruction, training, and main experiment. After reading the instructions, subjects went through 10 min of training (60 trials), where they were required to obtain at least 40% performance. The training comprised only stimuli of the two greatest change sizes (110%, 140%). Three subjects did not attain the criterion level of performance and were not tested further.

The main experiment was composed of two sessions of about 70 min each, comprising a total of 930 trials, corresponding to 60 repetitions of each condition. The two sessions were never more than 2 days apart. After reading the instructions, subjects were aware that the change could arise at any moment on each trial and that their task was to detect it within the 2 s window.

Visual feedback was always displayed on the screen in front of them; either a red square was displayed when the button was pressed before the change (false alarm), or when the button was not pressed within the 2 s time window after the change (miss). A green square was displayed when the button was pressed after the change but within the 2 s window.

In addition, sound level was roved from trial to trial; it was chosen randomly between 60 and 80 dB SPL (sound pressure level). This procedure is classically applied to prevent subjects from adopting any absolute level strategy. The inter-trial interval was ~1 s with a small, random jitter (<0.1 s) depending on computer load.

## 2.5 Data Analysis

We quantified the ability of the subjects to detect the change in stimulus statistics using two measures, hit rate and d-prime (d'). We also found reaction times to depend on the difficulty.

These measures were computed as a function of change size and change time. Since change times were distributed continuously but with an exponential distribution, the set of change times was binned with approximately exponentially increasing bin size (in order to achieve comparable numbers of trials in each bin).

To control for inattentive subjects, we set a 35% threshold for the total false alarm rate. Two subjects were discarded according to this criterion leaving a total of 10 subjects for the data analysis, with false alarm rates around 25%.

### 2.5.1 Hit Rate and Reaction Times

We computed a subject's hit rate as the fraction between successful detection (hits) out of the total trials for which the change occurred before the subject's response (hits + misses). False alarms were excluded from the hit rate computation, since they occurred before the subject was exposed to the change (see d' below for an inclusion of false alarms). We obtained reaction times by subtracting the change time from the response time in each trial.

### 2.5.2 d' Analysis

We computed d' values to assess the ability to detect changes, while taking their false alarm rate into account. Due to the present task structure, d' was computed as a function of time from stimulus onset (see Fig. 1e for an illustration), approximated as $d'(t) = Z(HR(t)) - Z(FAR(t))$, where Z(p) is the inverse of the cumulative Gaussian distribution. HR(t) is the hit rate as a function of time t since stimulus onset. Hit rate was computed as the fraction of correct change detections, in relation to the number of trials with changes occurring at t. Similarly, the false alarm rate FAR(t) was computed as the number of false alarms that occurred over all 2 s windows (starting at t), in which no change in statistics occurred. The window of 2 s was chosen to be compatible with the hit rates in the 2 s decision window. d' was computed separately for different times and change sizes, yielding only a limited number of trials per condition. To avoid degenerate cases (i.e. d' would be infinite for perfect scores), the analysis was not performed separately by subject, but over the pooled data. Confidence bounds (95%) were then estimated by grouping data from all subjects. The analysis was verified on surrogate data from a random responder (binomial with very low p at each point in time), providing d' very close to 0 on a comparable number of trials.

### 2.5.3 Hit rate Dynamics

In order to compare the hit rate dynamics for different change sizes, we fitted (least-square non-linear minimization) a cumulative Erlang distribution to the data according to:

$$P(\Delta_c, t_c) = P_0(\Delta_c) + P_{max}(\Delta_c) * \gamma(k, t_c / \tau(\Delta_c)) / (k-1)!$$

where $P_0$ is the minimal hit-rate, $P_{max}$ is the maximal hit rate, $t_c$ is change time, $\Delta_c$ the change size, $\gamma$ the incomplete gamma function, $\tau$ the function rate, and $k$ controls the function shape. $k$ was kept constant across subjects and change sizes, assuming the shape of the hit rate curves is invariant, which appeared to be the case in our sample.

### *2.6 Statistical Analysis*

In the statistical analysis only non-parametric tests were used. One-way analysis of variance were computed with Kruskal-Wallis' test, the two-way were computed using Friedman's test. Unless is indicated otherwise, error bars correspond to twice the standard error of the mean (SEM). All statistical analyses were performed using Matlab (The Mathworks, Natick).

## 3 Results

Human listeners ($n = 15$, 10 performed to criterion level) were presented with a continuous acoustic texture and required to detect a change in stimulus statistics, which could occur at a random time, after which they had 2 s to respond. Changes occurred in the probability of occurrence of tones in randomly chosen spectral regions (Fig. 1a).

Several parameters of the change could determine its salience, such as its size, timing. These parameters were tested in a single experiment with ~1000 trials over two sessions per subject.

### *3.1 Detection of Changes in Statistics is Consistent with Integration*

The ability of subjects to detect a change in stimulus statistics (Fig. 1a) improved with the time they were allowed to listen to the initial statistics of the texture (Fig. 1d, progression along ordinate, performance measured as hit rate, $p < 10^{-6}$, Friedman test). Hit rate increased monotonically to an asymptotic level for all change sizes (4 levels, [50, 80, 110, 140] %). Asymptotic hit rate depended on the change size, with bigger changes in marginal probability leading to greater asymptotic hit rate (Fig. 1d, different colours). Although asymptotic hit rate was above chance (26.6 %) for all change sizes, the increase with change size was large and significant (from 50 to 95 %, $p < 10^{-7}$, Friedman test).

Change size also influenced the shape of the dependence on change time, such that greater change sizes lead to improved hit rate already for shorter change times

(Fig. 1d). This translates to a combined steepening and leftward shift of the hit rate curves with change size. Significance of this effect was assessed by fitting sigmoidal hit rate curves of individual subjects with a parametric function (see *Methods*) in order to extract the change size-dependent time constant (Fig. 1f). Hit rate time constants $\tau$ significantly decreased with respect to change sizes ($p < 10^{-6}$, Friedman test).

The results above stay true if false alarms are taken into account (Fig. 1h). For this purpose, we computed a time-dependent d' measure (see Fig. 1g for an illustration, and *Methods*). Classical d' analysis does not apply in the present dynamic decision paradigm, since the eventual outcome of a trial is not accessible at early times (i.e. when computing d'(0.5 s), a trial with a hit at 5 s should not be counted as a hit, but as a 'correct reject up to 0.5 s; but when computing d'(4 s), the same trial will be counted as a hit). In accordance with the perceived difficulty of the detection task, d' values are significantly positive (Fig. 1h, based on the 95 % confidence bounds) and increase with change time (Friedman test, $p < 0.01$), but stay relatively low especially for 110 % and below, indicates how challenging the task was, leading to a substantial number of false alarms.

The above results are consistent with predictions from statistical integration of information. First, unsurprisingly, change detection should improve with greater change sizes. Second, change detection should improve, but saturate, as the estimation of the statistics of the first stimulus converges, i.e. if longer observation time is available. Third, change detection of bigger changes should require less precision in the estimation, translating to higher hit rate already for shorter observation times.

### *3.2 Reaction Times are Consistent with Integration*

The dependence of performance on change time suggests a dynamical integration mechanism to play a role in the decision process. While the dependence on change time indicates an ongoing estimation of the initial statistics, it does not provide much insight into the estimation of the post-change statistics. We hypothesized that reaction times here could be informative. Reaction times had previously been demonstrated to depend on task difficulty (Kiani et al. 2014), and difficulty in the present paradigm correlates inversely with the availability of a greater difference in change size, i.e. statistics (Fig. 2a).

Reaction time distributions changed both in time and shape as a function of change size (Fig. 2b). Median reaction time correlated negatively with change sizes ($p < 0.001$; Fig. 2c), in a manner that corresponded to the inverse of the increase in hit rate with larger change sizes. In addition, the onset time of the distributions decreased with change size, together suggesting that certainty was reached earlier for bigger step-sizes.

As a function of change time, the reaction time distribution changed in a qualitatively similar manner, however, less pronounced (Fig. 2e). Median reaction times correlated negatively with change times, mirroring dependence of hit rate on change

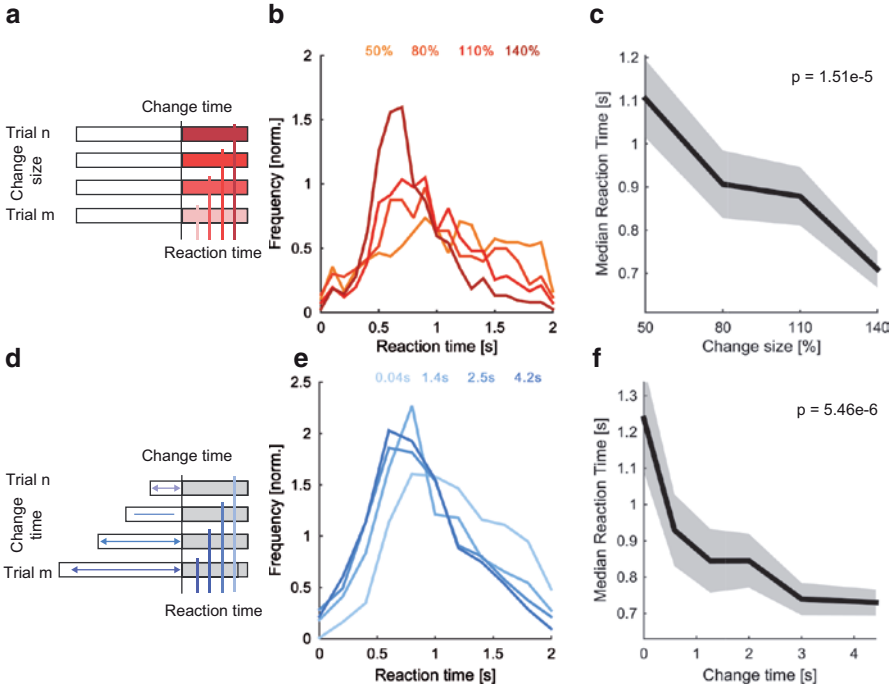


**Fig. 2** Reaction times reflect task difficulty and temporal integration. **a**–**c** If reaction times are separated by change size (**a**), one observes a progression in the shape (**b**) as well as median response time (**c**). The shape of the reaction time distribution sharpens with increasing step-size and the median response time decreases significantly ($p = 1.51e\text{-}05$, Kruskal-Wallis). **d**–**f** If reaction times are separated by the duration of the change time before the change in statistics occurred (**d**), the distribution of reaction times again sharpens with longer change times (**e**) and the median reaction time (**f**) decreases significantly ($p = 5.46e\text{-}06$, Kruskal-Wallis)

times ($p < 0.001$; Fig. 2f). This dependence can already be seen in the raw data (Fig. 1c), where hit trials (black) for longer change times exhibit shorter reaction times. Again the onset time decreased correspondingly with longer change time, suggesting more accurate estimation of the initial statistics.

## 4 Discussion

In the present study we used a simplified acoustic texture to study detecting changes in the statistics of complex acoustic stimuli. The results suggest that humans are able to integrate stimulus statistics, as expressed in their performance and reaction times depending on the parameters of the task.

## 4.1 Dynamic Representation of Spectral Statistics

The stimulus of the present paradigm can be considered a stochastic version of the appearance or increase in loudness of a sound source within a natural auditory scene. Its design is a compromise between complexity of the spectrotemporal structure and simplicity in specification since the parameter-space is reduced to a minimum and parameters are straightforward to interpret.

The present paradigm builds on research from the spectral and the temporal domain, e.g. profile analysis (Green and Berg 1991) or some paradigms in mismatch negativity. For example in profile analysis subjects assess the spectral shape of stimuli. In this context it has also been demonstrated that subjects can compare spectra, however, the results were not interpretable in a statistical sense, due to several reasons. First, the time scale of presentation is much longer in the present paradigm than typically in profile analysis, where the stimulus is presented for multiple seconds rather than subsecond presentations. Second, in the present context the stimulus is composed of randomly timed occurrences of tones in certain frequency bins (which defines the spectral profile over time), rather than static profiles from constantly presented tones or filtered noise. Third, the comparison between different profiles is performed dynamically, rather than in a two stimulus comparison task. As described above this allowed us to study the dynamic integration of stochastic information, and enables the collection of reaction times as another indicator of the underlying processes.

## 4.2 Future Directions

The present study lays some groundwork for future studies, by investigating some of the basic, previously unaddressed questions related to the integration of statistics in complex auditory stimuli, such as auditory textures. Further studies are required to address variations of the present paradigm, such as reductions in probability and continuous stimulus presentation. Attention was not controlled for in the present setting, although subjects were instructed to attend to the stimulus. Some subjects reported improved detection of change, when *not* attending to the stimulus, suggesting that a dual-task paradigm would be interesting where attention can be assessed and controlled.

**Acknowledgments** Funding was provided through the Advanced ERC ADAM 295603, the ANR-10-LABX-0087-IEC and the ANR-10-IDEX-0001-02-PSL*.

## References

Constantino FC, Pinggera L, Paranamana S, Kashino M, Chait M (2012) Detection of appearing and disappearing objects in complex acoustic scenes. PLoS ONE. doi:10.1371/journal.pone.0046167

Green DM, Berg BG (1991) Spectral weights and the profile bowl. Q J Exp Psychol Sect A 43(3):449–458

Kiani R, Corthell L, Shadlen MN (2014) Choice certainty is informed by both evidence and decision time. Neuron 84(6):1329–1342

McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron 71(5):926–940

McDermott JH, Schemitsch M, Simoncelli EP (2013) Summary statistics in auditory perception. Nat Neurosci 16(4):493–498