Aberrant perceptual judgements on speech-relevant acoustic features in hallucination-prone individuals

Julia Erb[1], Jens Kreitewolf[1], Ana P. Pinheiro[3], & Jonas Obleser[1,2]

[1] Department of Psychology, University of Lübeck, Lübeck, Germany
[2] Center for Brain, Behavior, and Metabolism, University of Lübeck, Lübeck, Germany
[3] Faculdade de Psicologia, Universidade de Lisboa (FPUL), Lisbon, Portugal

*Correspondence: Julia Erb, Maria-Goeppert-Str. 9a, 23562 Lübeck, Germany
Phone: +49 451 3101 3631, e-mail: julia.erb@uni-luebeck.de

Abstract

Hallucinations constitute an intriguing model of how percepts are generated and how perception can fail. Here, we investigate the hypothesis that an altered perceptual weighting of the spectro-temporal modulations which characterize speech contributes to the emergence of auditory verbal hallucinations. Healthy adults (N=168) varying in their predisposition for hallucinations had to choose the ‘more speech-like’ of two presented ambiguous sound textures and give a confidence judgement. Using psychophysical reverse correlation, we quantified the contribution of different acoustic features to a listener’s perceptual decisions. Higher hallucination proneness covaried with lower perceptual weighting of speech-typical, low-frequency acoustic energy. Remarkably, higher confidence judgements in single trials depended not only on acoustic evidence but also on an individual’s hallucination proneness and schizotypy score. In line with an account of altered perceptual priors and differential weighting of sensory evidence, these results show that hallucination-prone individuals exhibit qualitative and quantitative changes in their perception of the modulations typical for speech.

Key words: Speech perception, psychoacoustics, reverse-correlation, spectro-temporal modulations, auditory verbal hallucinations, schizotypy

Introduction

A major challenge of sensory neuroscience remains to understand how adaptive top-down weighting of sensory evidence due to, e.g., ongoing task demands influence percepts. As hallucinations occur in the absence of an external stimulus, they constitute an intriguing model for the generation of percepts. Hallucinatory experiences, mostly visual or auditory, are prevalent in psychotic disorders such as schizophrenia, but also have an estimated prevalence of 6-13% in the general population (‘non-clinical voice hearers’) [1], consistent with the hypothesis that psychosis exists on a continuum with normal experience [2].

Prior expectations are suggested to be critical in the generation of hallucinations. Evidence suggests both overly strong or weak priors in hallucination proneness and psychosis [3]. For example, prior knowledge of an image leads to an advantage in recognizing that image when it is degraded; individuals at risk of psychosis are more susceptible to this perceptual advantage [4]. A recent sensory conditioning study [5] used a visual conditioned cue to predict a faint auditory stimulus presented at threshold. All participants experienced conditioned auditory hallucinations when presented with the visual (but not the auditory) stimulus. However, individuals with hallucinations were more susceptible to such conditioned hallucinations [5]. Non-clinical voice hearers listening to degraded (sine-wave) speech, also show stronger expectations to hear speech than controls who recognize the presence of speech later [6]. Such observations imply an increased bias towards top-down information in hallucination proneness.

Regarding the neurobiology of auditory verbal hallucinations (AVH), functional MRI studies report activation of auditory cortex [7], including primary auditory cortex [8], when patients experience AVH (for review see [9]). However, it is unclear in how far the general response properties of auditory cortex are altered in voice hearers. A recent strain of research provides converging evidence that the healthy human auditory cortex analyzes sounds along so-called spectro-temporal modulations: The auditory pathway is thought to not only implement forms of “tonotopic” frequency analysis[10], but to rather represent sound as frequency-specific spectral and temporal modulation filters ([11,12] for neurobiological evidence see e.g., [13,14]). Hallucinations in schizophrenia have been linked to deficits in object formation [15]. Auditory object formation in turn is known to rely heavily on the extraction of the spectro-temporal modulations in the auditory scene [16].

In the present study we therefore examine whether differences in the perception of these spectro-temporal modulations abundant in speech [17] can be linked to a propensity towards auditory hallucinations. We here investigate processing of spectro-temporal modulations in individuals presenting with varying, non-clinical degrees of predisposition to hallucinations.

First, we establish individual listeners' "*speechiness* kernels", that is, an individual template of those acoustic features that elicit a speech percept. To this end, we present two ambiguous sound textures in noise [18] to human listeners and ask them to the 'more speech-like' one. This allows us to retrieve their internal representation that drives the categorization into speech, using the psychophysical technique of reverse correlation [19,20]. Second, we relate those speechiness kernels to the degree of individual schizotypal traits (unusual perceptual experiences subscale [21,22]), and to individual hallucination proneness [23,24], probing a psychosis continuum. Compared to studies with psychotic patients, the study of non-clinical participants has the important advantage to circumvent confounding factors such as medication and presence of other symptoms (e.g., negative symptoms [25]).

The results pose an intriguing link between current models in computational psychiatry and recent advances in modelling the perceptual and neural response in auditory neuroscience.

Results

In a short online (*N* = 131) and an extended lab version (*N* = 37) of a 2-AFC experiment with confidence judgement, participants had to choose the 'more speech-like' of two presented ambiguous sound textures (Fig. 1). Using reverse correlation, we obtained perceptive fields termed "*speechiness* kernels" that quantify the contribution of different acoustic features to a listener's perceptual decisions. Hallucination proneness was assessed with the Launay-Slade hallucination scale (LSHS). Additionally, we evaluated schizotypy using the Schizotypal Personality Questionnaire (SPQ) with a particular interest in the subscale "unusual perceptual experiences" (SPQ-UP) as second measure of predisposition to unusual perception.

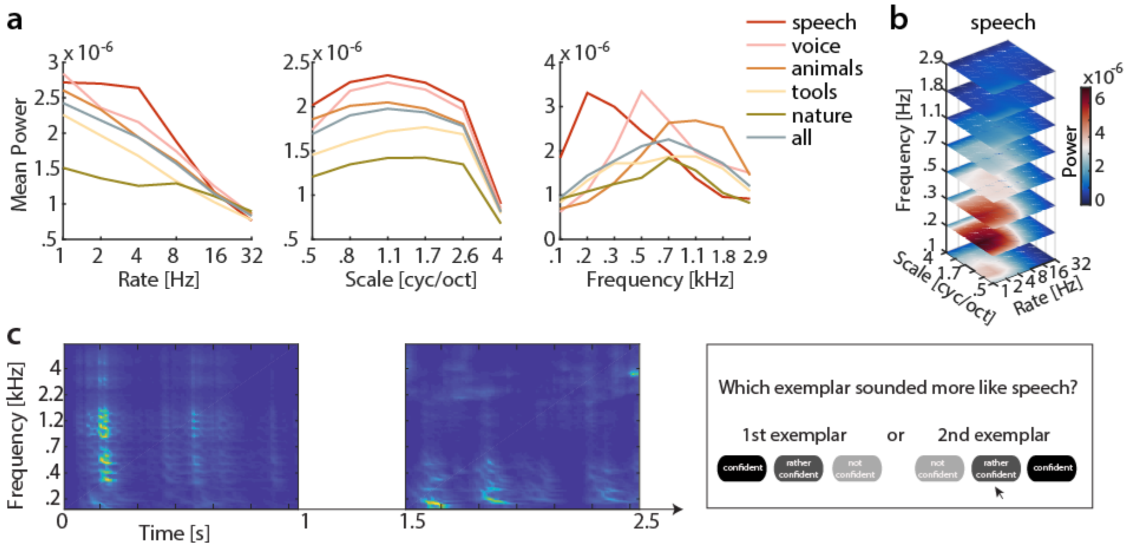

Figure 1. Stimuli and task. (a) Modulation spectra for sound textures, marginalized for each acoustic dimension (rate, scale, frequency) and split by categories of the original sounds from which the textures were synthesized (see legend). (b) Average modulation spectrum for the speech textures. (c) In a 2-alternative-forced-choice (2-AFC) task with confidence judgement, participants were presented with two sound textures (here shown as spectrograms) and were asked to simultaneously express their decision and confidence about which exemplar sounded more like speech.

Schizotypal traits and hallucination predisposition

In the lab experiment, LSHS scores ranged from 1 to 32 (median = 10; max possible score 48) and global SPQ scores ranged from 3 to 37 (median = 16, max possible score 74). In the online experiment, LSHS scores ranged from 1 to 33 (median = 9) and global SPQ scores ranged from 0 to 41 (median = 15; see Fig. 2a).

Across both experiments (*N* = 168), LSHS scores were uncorrelated with age (*r* = –.098, *p* = .206) and gender (*r* = –.077, *p* = .321). Similarly, global SPQ scores were essentially uncorrelated with age (*r* = .099, *p* = .201) and gender (*r* = .140, *p* = .070). More importantly, LSHS scores, global SPQ scores and the subscale SPQ-UP were all substantially correlated amongst each other (Fig. 2b). The intercorrelations of 18–42 % shared variance emphasize the convergent validity of the questionnaires used to measure individual predisposition to unusual perceptual experiences here.

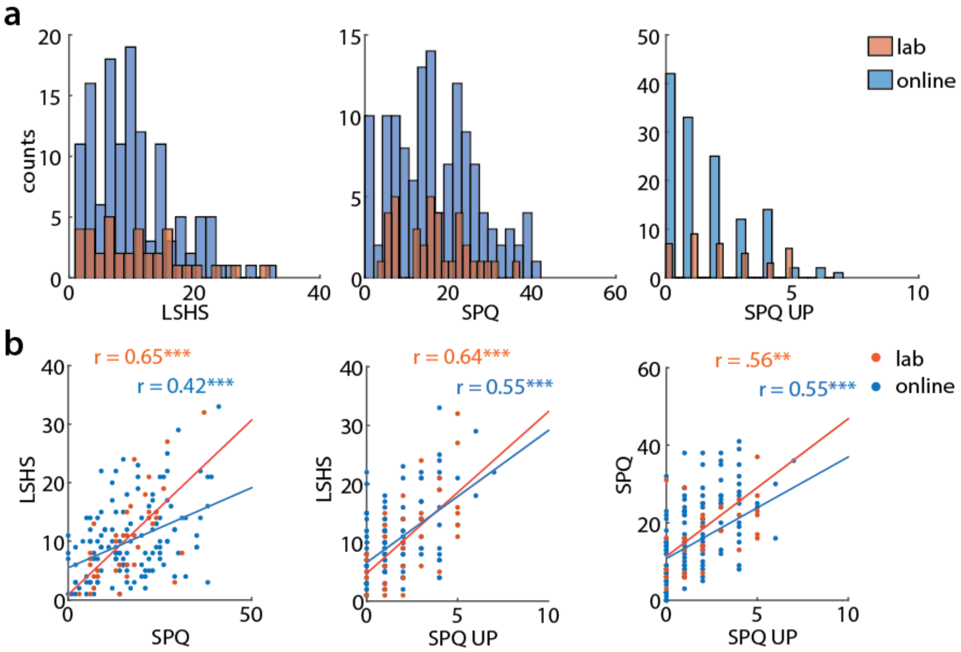

Figure 2. Schizotypy and hallucination scale results. (a) Histograms for LSHS and SPQ scores separately for lab (blue) and online (orange) experiment. Note that the maximal possible score is 42 for LSHS, 74 for SPQ, and 9 for the subscale SPQ-UP. (b) Scatter plots showing the correlations between LSHS, SPQ total score and SPQ-UP. Pearson's correlation coefficients are shown separately for lab (red) and online experiment (blue). All scales are substantially correlated, R^2 ranges from .18 to .42. LSHS: Launay-Slade hallucination scale; SPQ: schizotypal personality questionnaire; SPQ-UP: subscale unusual perceptual experiences. ** $p < .005$, *** $p < .001$.

Speechiness kernels

First, we analyzed how participants' judgements of speechiness varied as a function of the spectro-temporal modulations contained in the sound textures. To obtain such speechiness kernels, we used reverse correlation by contrasting the averaged spectro-temporal modulations of the stimuli judged as more versus less speech-like. Speechiness kernels averaged across participants were highly correlated between the lab and online studies (Fig. 3a). The marginal profiles for temporal modulations peaked at ~4 Hz and for frequency at 200 Hz, indicating high speechiness judgements when acoustic power was high at slow temporal modulations and low frequencies (Fig. 3b,c). As an outlook, those peaks were driven by the trials on which the participants were confident (see Fig. 6a). We permuted participants' responses (n = 10,000 permutations) to obtain the empirical null distribution of the speechiness kernels (Fig. 3b, yellow line).

We *z*-scored empirical kernels relative to the null distribution; *z*-scores proved significant (i.e., $|z| > 1.96$) for temporal rates of 4–8 Hz in the lab experiment, and for low and high frequencies in both the lab and online experiment (Fig. 3c).

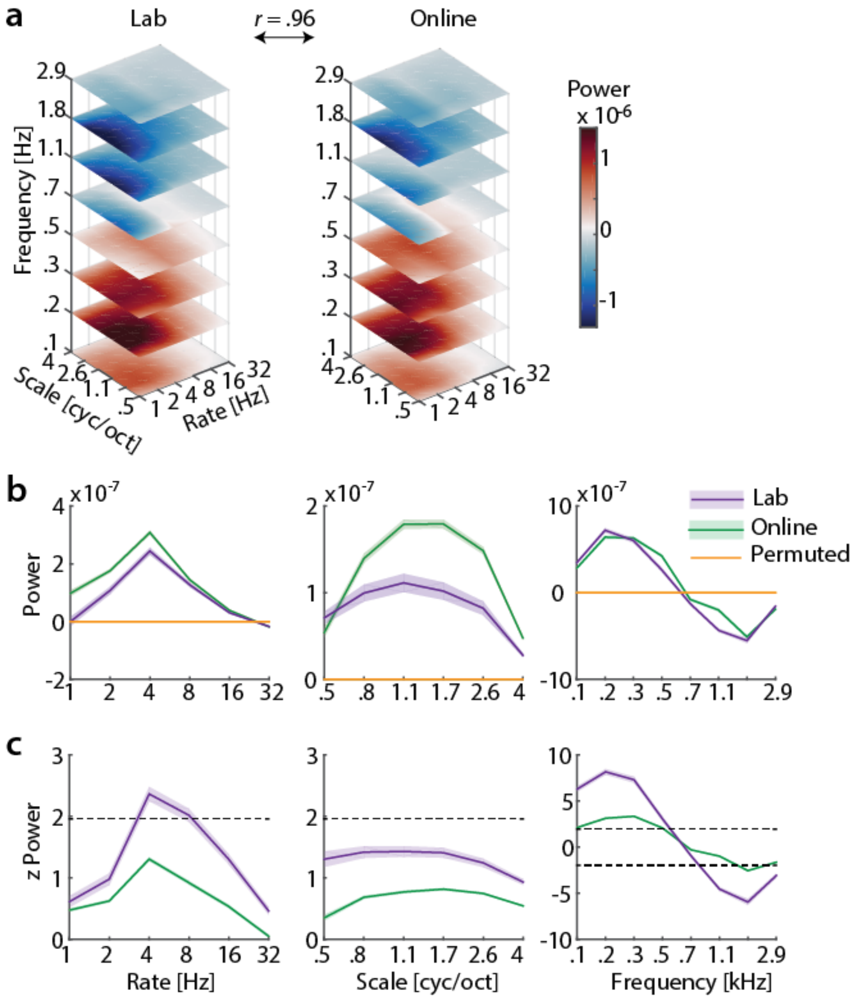

Figure 3. Speechiness kernels. (a) Averaged speechiness kernels for lab and online study are highly correlated (Pearson's *r*). (b,c) Marginal profiles of speechiness kernels (mean ± standard error) for lab and online study separately before (b) and after (c) z-scoring relative to the empirical null distribution (obtained with n = 10,000 permutations). Absolute z-scores exceeding 1.96 are considered significant (indicated by the dashed line).

Kernel stability. In the lab experiment, average speechiness kernels based on the first 100 trials were highly similar to the total kernel (i.e., comprising all 540 trials, mean [SE] Pearson's r = 0.926 [0.026], see supplementary Fig. S1). We took this finding as evidence that the lower number of 108 trials used in the online experiment suffice to obtain a stable estimate of the speechiness kernel.

Relation of acoustic feature weighting to hallucinatory predispositions

Next, we asked whether the individual extent of hallucination proneness and schizotypy is related to features of the speechiness kernel. First, we conducted a principal component analysis (PCA) to reduce the dimensionality of the speechiness kernels from 288 components (one for each feature of the speechiness kernel) to the first six components (see Methods). This procedure was justified by the drop in eigenvalues after the sixth component (see scree plot Fig. 4a). There was one clear dominant component (component 1) and two minor components (component 2 and 3; whose eigenvalue still clearly exceeded the eigenvalues of surrogate data, Fig. 4a).

Both the first and third component were characterized by high acoustic energy at high frequencies (Fig. 4b, c). The PCA approach also held the advantage of yielding, by design, independent regressors to be subsequently used in a linear model. We used individual component scores of the first six components to predict individuals' LSHS and SPQ-UP scores in a multiple regression analysis (Table 1).

Table 1: Multiple Linear Regression predicting LSHS and SPQ UP scores based on the first six principal component (PC) scores of the individual speechiness kernels based on data from N = 168 individuals. Shown are beta estimates of effects of interest with a 95% confidence interval (CI).

	LSHS score			**SPQ UP score**		
Predictors	ß	95% CI	*p*	ß	95% CI	*p*
PC No.1	0.03	-0.12 – 0.18	0.689	-0.01	-0.15 – 0.15	0.958
PC No.2	-0.05	-0.20 – 0.10	0.529	-0.11	-0.26 – 0.03	0.129
PC No.3	**0.19**	**0.04 – 0.34**	**0.013**	**0.20**	**0.05 – 0.34**	**0.011**
PC No.4	-0.05	-0.20 – 0.10	0.520	-0.14	-0.29 – 0.01	0.065
PC No.5	0.10	-0.05 – 0.25	0.194	-0.02	-0.17 – 0.13	0.763
PC No.6	-0.05	-0.20 – 0.10	0.485	-0.09	-0.24 – 0.06	0.215
R^2	0.056			0.080		

The third component of the high-dimensional acoustic speechiness kernel, which was dominated by high-frequency energy content (Fig 4b, c), covaried significantly with an individual's tendency towards aberrant perception, that is, both the LSHS (Pearson's r = 0.192, p = 0.013, Bayes factor BF_{10} = 2.126) and SPQ-UP score (r = 0.195, p = 0.011, BF_{10} = 2.306; Fig. 4d; see also Table 1). These results provide evidence for an association of higher hallucination proneness and schizotypal traits (i.e, unusual perceptual experiences) with classifying stimuli into speech that are characterized by high frequency components (Fig. 4b, c). This effect is illustrated in the markedly different speechiness kernels of the two individuals with the most extreme scores for the third component (Fig. 4e).

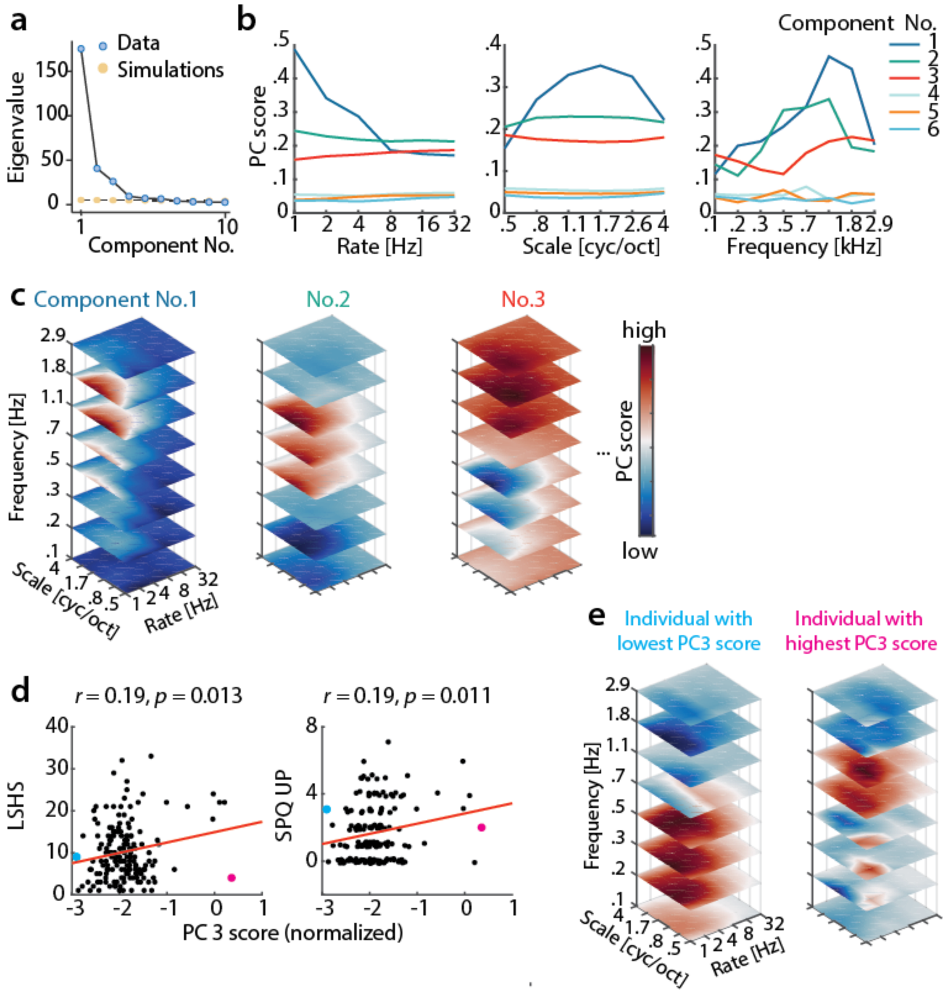

Figure 4. Principal component analysis (PCA) of the speechiness kernels jointly from lab and online experiment. (a) Scree plot for the PCA. For subsequent analyses, only the first six components were retained because the eigenvalue (blue dots) for all further components was smaller than randomly generated eigenvalues (yellow dots). There was one clear dominant component (component 1) and two minor components (component 2 and 3). (b) Marginal profiles of the first six components. (c) First three components in the modulation space. (d) Scatter plot of the scores of the third component as a function of hallucination proneness and schizotypal traits. LSHS and SPQ-UP scores are correlated to scores for component no. 3 only. To jitter the integer scores slightly for display purposes only, a uniformly distributed random quantity between –0.15 and +0.15 was added to the SPQ-UP

scores. (e) For visualization purposes, we show example speechiness kernels for two individuals with extreme scores for component no. 3: the lowest (left; cyan dot in the scatter plots in d) and highest PC3 scores (right; magenta dot in the scatters plot in d). PC: Principal component; LSHS: Launay-Slade hallucination scale; SPQ-UP: schizotypal personality questionnaire subscale unusual perceptual experiences.

Sound discriminability

In the current experimental setup, it is to be expected that the speech-like modulation content in a presented pair of sound textures (i.e., the sensory evidence) should be a prominent driver of all listeners' speechiness judgements.

Under the assumption that the speechiness kernels reflect individual perceptive fields of speech, we tested the effects of these perceptive fields on sound discriminability. We calculated pairwise distances between sounds based on their original modulation representations as well as on the representations obtained by weighting the original representations by the individual speechiness kernels. Then, for each sound separately, the features were normalized between their minimum and maximum (effectively scaling them between 0 and 1). For each trial, sound pair distance was calculated as the Euclidian distance between each of the 288 features. A comparison between original and projected sound pair distance showed that projecting through individual speechiness kernels increased discriminability of sound pairs (Fig. 5a).

To quantify benefits from individual speechiness kernels, we fitted individual linear slopes to projected as a function of originally presented sound-pair distance. Individual slopes ("speechiness kernel benefit") were significantly higher than one ($t(167) = 32.61$, $p < .001$), indicating the "warping" of an acoustic into a perceptual distance representation and validating the present perceptual speech–kernel approach. Notably, however, individual speechiness kernel benefits were unrelated to LSHS scores (Fig. 5b) with evidence for the absence of an effect as indicated by the Bayes Factor ($BF_{01} = 8.749$). These results suggest that internal templates for speech amplify the discriminability of sound textures but are formed independently from an individual listener's hallucination proneness.

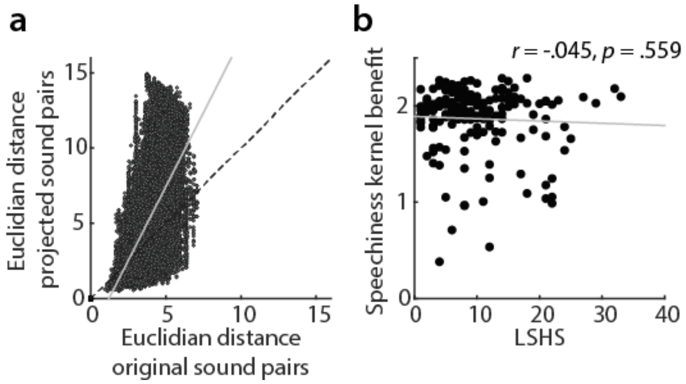

Figure 5. Effect of speechiness kernels on discriminability of sound pairs. (a) We filtered all sounds (presented across experiments and participants, *n* = 34,128 sound pairs) by individual speechiness kernels, leading overall to higher discriminability (euclidian distance in the modulation space) of projected compared to original sound pairs. (b) In this space, we fitted individual linear regression lines. The slope of this linear fit ('speechiness kernel benefit') was above 1 for most participants, indicating that filtering with individual speechiness kernels improved discriminability of sound pairs. However, this benefit was unrelated to hallucination proneness, as shown by Pearson's *r*. LSHS: Launay-Slade hallucination scale.

Confidence judgements

To investigate whether confidence affects the speechiness kernels, we calculated three different kernels, one for each confidence level separately (Fig. 6a). Although the shape of the kernels was similar for all confidence levels, higher confidence amplified the magnitude of the kernel (see also Fig. 6b). The proportion of "confident" responses was positively correlated both with the LSHS scores and global SPQ scores, indicating that participants with higher hallucination proneness and higher schizotypy reported more often that they were confident about their speechiness judgements (Fig. 6d).

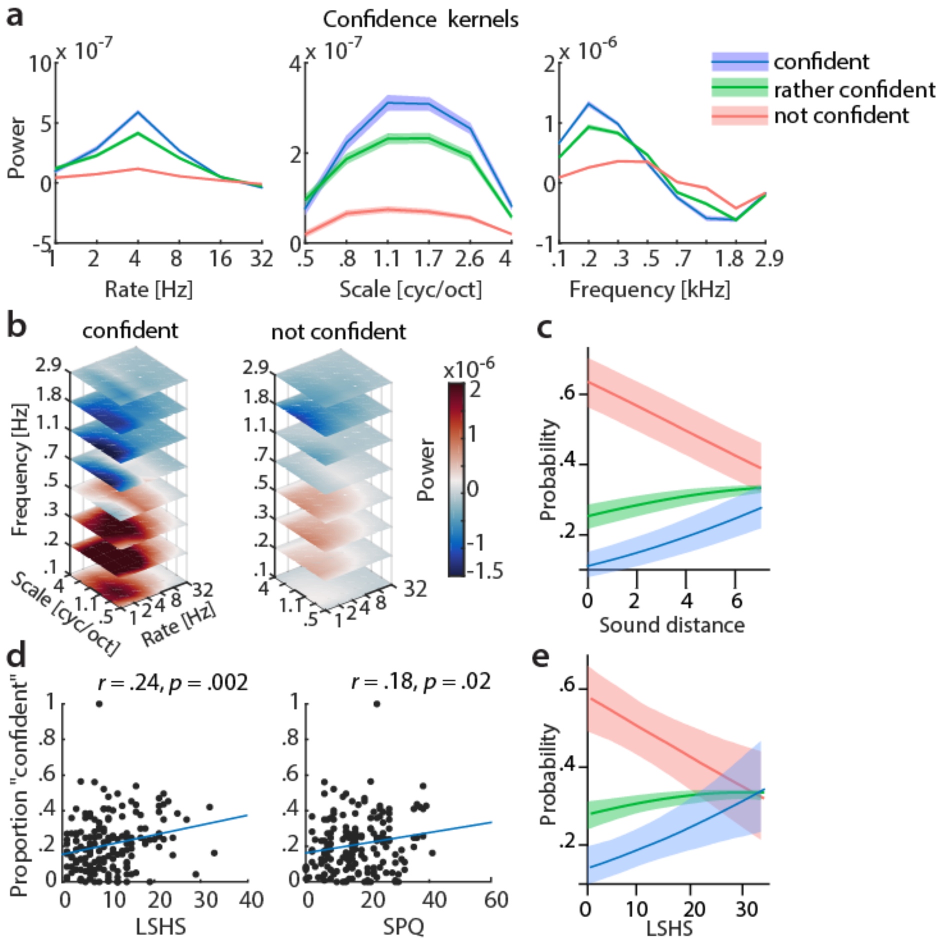

Figure 6. Confidence judgements in speechiness kernels (jointly from lab and online experiment) and relation of hallucination proneness to confidence judgements. (a) Marginal profiles of speechiness kernels (mean ± SE) for three different confidence levels separately. (b) Speechiness kernels for the two extreme confidence levels (confident, not confident). (c) Estimates from a generalized linear mixed model (GLMM) predicting confidence levels based on sound pair distance; the model also has the predictors LSHS, SPQ and experiment type (see below); same color legend same as in (a). (d) Proportion of "confident" judgements correlated (Pearson's *r*) with LSHS and global SPQ scores. (e) Estimates from a GLMM predicting confidence levels based on LSHS, model also has the predictors sound distance, SPQ and experiment type (see below); same color legend same as in (a). LSHS: Launay-Slade hallucination scale, SPQ: schizotypal personality questionnaire.

A more mechanistic explanation of this correlation would afford that this relation of confidence judgements to hallucination proneness also holds at the trial-by-trial level, where we can account for stimulus discriminability, experiment type, and subject-specific intercepts.

Using ordinal linear mixed-effects regression, we regressed trial-by-trial confidence judgements (on a Likert scale from 1 [“not confident”] to 3 [“confident”]) against the following predictors: trial-wise sound pair distance (Euclidian distance in the modulation representation of the two sounds presented on a given trial); the personality variables LSHS score and global SPQ score; as well as experiment (a binary indicator variable coding online versus lab), and a subject-specific random intercept.

Greater sound-pair distance, but also higher LSHS and global SPQ scores were significantly associated with higher confidence judgements (Table 2; Fig. 6 c,e). The interactions of sound pair distance with LSHS and global SPQ score, respectively, were not significant predictors of confidence judgement when considering a 95% credible interval, indicating that the confidence judgement based on sensory evidence did not depend substantially on hallucination proneness. Observing these data was about thirteen times more likely under a model including the LSHS score than under a null model with the same parameters except LSHS, as evidenced by an average Bayes Factor $BF_{LSHS–null}$ of 13.31, 95% CI [12.56; 14.05]. Echoing the convergent validity of LSHS and SPQ, global SPQ score also proved a significant predictor of confidence judgement: a very comparable magnitude was observed for the Bayes Factor of a model including the global SPQ score instead ($BF_{SPQ–null}$ = 13.21, 95% CI [12.35;14.07]).

Table 2: **Generalized Linear Model** (Ordinal regression) predicting confidence judgements on a Likert scale from 1 (“not confident”) to 3 (“confident”). Shown are estimates of effects of interest as Log Odds with a 95% Bayesian highest posterior density interval (labelled “95% CI”, credible interval). The model entailed data from a total of 34,128 single trials from N=168 participants.

	Confidence Judgement	
Predictor	Log Odds	95 % CI
Sound pair distance	**0.09**	**0.08 – 0.11**
LSHS score	**0.12**	**0.02 – 0.22**
Global SPQ score	**0.11**	**0.02 – 0.19**
Experiment (lab [0] vs. online [1])	–0.12	–0.32 – 0.08
Sound distance × LSHS score *	–0.02	–0.03 – 0
Sound distance × global SPQ score	0.00	–0.01 – 0.02
Sound distance × Experiment *	–0.03	–0.06 – 0

** Significant for a 90% CI interval*

Discussion

In how far does hallucination proneness in non-clinical participants manifest in the aberrant perceptual judgement of acoustic features, namely spectro-temporal modulations? We studied this using a simple “speechiness” judgement with confidence ratings based on synthesized sound textures, both in a short online and an extended lab experiment and gathered a total of N=168 data sets.

First, we found individuals’ scores on both the schizotypy personality questionnaire (SPQ) subscale on ‘unusual perceptual experiences’ and on the Launay–Slade Hallucinations Scale (LSHS) to covary with the degree to which they classified textures as ‘speech’ that were lacking the speech-typical low-frequency dominance.

Second, those individuals scoring higher on either of these scales were more confident in their perceptual decisions. Trial-wise confidence judgements were expectedly driven by acoustic stimulus distances (i.e., the sensory evidence available), but also—to an equal magnitude—by LSHS scores thought to capture hallucination proneness (i.e., a perceptual prior or predisposition, see Fig. 7).

The present results are remarkably in line with an account of altered perceptual priors and a differential weighing of sensory evidence in hallucination-prone individuals, with accordingly changed perceptual decisions when classifying speech-like sounds.

Speech perception in Hallucination proneness

The PCA analyses allowed us to examine to which degree different acoustic features were used for perceptual decision-making in varying levels of hallucination proneness. We found the presence of high-frequency components and – maybe more importantly – the absence of the speech-typical low temporal modulations to contribute to the classification into speech in hallucination-prone individuals.

Hallucination proneness has been associated with more false alarms in auditory signal detection tasks. An increased false alarm rate was observed in a tone detection task with a conditioning visual stimulus in hallucinating individuals with and without psychosis [5]. In a speech in noise detection study, participants were to indicate “whether they had heard a voice”: Non-clinical hallucination-prone adults had more false alarms and expressed a liberal response bias [26].Yet, this study could not disentangle whether distinct acoustic features contribute to a bias towards a speech percept.

Why do the speechiness kernels in hallucination proneness exhibit particular high frequency dominance while being flat in the other two dimensions (i.e., the temporal and spectral modulations)? Our findings of atypical speech perception parallel the recent observations of differences in speech production in schizophrenia: aberrant acoustic patterns of vocal expressions (e.g., in pitch variability) have been reported in schizophrenic patients [27]. The underlying cause common to aberrant speech perception and production may be a deficit in auditory object formation as postulated for schizophrenia[15], consistent with the notion that object formation relies on the intact extraction of spectro-temporal cues[16].

However, sound distances on a given trial (akin to the available sensory evidence) were similarly amplified by individual speechiness kernels, independent of hallucination proneness. Perceptive (from psychophysics) and receptive fields (i.e., from physiological experiments) obtained using reverse correlation can be surprisingly similar [28]. Under the premise that the speechiness kernels reflect individual perceptive fields, this finding indicates that although speech-relevant modulations may be encoded differently by hallucination-prone individuals, this may not readily be reflected in their discriminability performance. Rather, we found differences in confidence judgements to depend on hallucination proneness and schizotypy.

Confidence Judgements in Hallucination proneness

In a framework of Bayesian models of perception, the current results pose new evidence for changes to prior expectations in hallucination proneness [29,30]: The data support an account of an increased precision or decreased variance in individual perceptual priors [31]. The statistical model of single-trial confidence judgements in their own choices ("which [sound] is the more speech-like one?") here provides important evidence.

First, as expected, the sensory evidence available on a given trial (i.e., width of an internal likelihood representation) exerts an impact on confidence. The expressed confidence can be thought of as the width or inverse precision of the posterior, reflecting the "noise" or uncertainty in one's perceptual judgement [31]. Accordingly, individual trait-like predispositions to perceive hallucinations can be considered stronger (i.e., less noisy and variable) perceptual priors, and they should hence contribute to a stronger confidence (i.e., smaller width of the posterior) in one's own perceptual judgements. While this has been a guiding conjecture in the field of computational psychiatry (e.g., [32]), it is borne out by the present data using an auditory reverse-correlation analysis technique. Figure 7 provides a schematic illustration of the evidence provided within a Bayesian framework. In sum, overly precise or strong priors that have previously been claimed for psychosis and hallucinations [3] may contribute to an aberrant percept of 'speechiness'.

The present results add novel evidence to a strain of findings supporting that prior beliefs mediate hallucinations. In the sensory conditioning study by Powers, et al. [5], individuals who hallucinate experienced more often conditioned auditory hallucinations when presented solely with the visual conditioning stimulus. In a Hierarchical Gaussian Filter [33], this expressed as a higher weighting of prior beliefs over sensory evidence in hallucinating individuals. Psychotic participants were less likely to update their prior beliefs when presented with new evidence, as shown by their smaller volatility estimate [5]. Similarly, in a visual degraded image recognition task, psychotic patients or individuals at risk of psychosis favored top-down prior knowledge of an image over available sensory evidence [4].

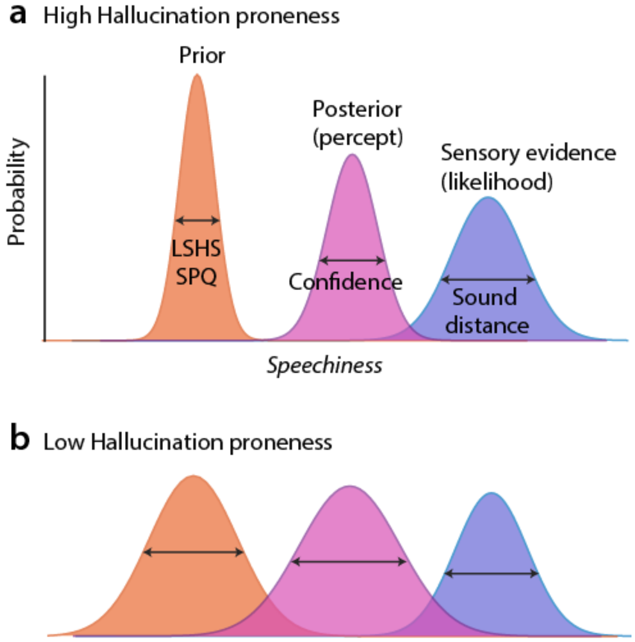

Figure 7. The current results in a framework of Bayesian models of perception. Perceptual inference in terms of the prior, likelihood and posterior (percept) is represented by Gaussian distributions over a perceptual dimension of 'speechiness'; their widths represent precision. Both hallucination proneness and schizotypy (LSHS and SPQ) are thought to be proportional to the inverse width of the prior, while confidence being proportional to the inverse width of the posterior. Sensory evidence was operationalized as Euclidian sound distance. (a) In high hallucination-prone participants, the prior precision for a percept of 'speechiness' is thought to be higher ('stronger prior'), contributing to the observed stronger confidence judgements (i.e., posterior precision) as compared to (b) the low hallucination-prone.

An important open question remains at which neuronal levels these aberrant percepts are computed. In contrast to tinnitus where usually simple tones or noises are perceived, AVH relate to the perception of complex sounds, that is, voices, in the absence of an external source. Similar to AVH [3], tinnitus has been proposed to be rooted in overly precise sensory evidence [34]. In tinnitus, layer-specific effects in A1 have been postulated to lead to a sharpening of a weak prior [34].

Beyond auditory cortex [9], AVH have been linked to changes at higher-level computations. For example, non-clinical voice hearers have been found to rely more heavily on an executive attention network including cingulo-opercular and frontal cortex when listening to degraded ("sine-wave") speech [6]. Future studies are expected to reveal whether the general response properties of auditory cortex are altered in hallucination proneness. The current study opens a specific and promising avenue using validated auditory perceptual-filtering models [11,12,17]: Spectro-temporal modulations currently form a core tenet of auditory neuroscience, tractable for non-human animal research (e.g.,[35]) as well as human functional neuroimaging (e.g., [36])

Conclusions

In sum, the current results endorse a continuum hypothesis of psychosis [2] showing that individuals with different degrees of schizotypy – who sometimes experience auditory hallucinations but are not diagnosed with any psychotic disorder – do have distinct signatures of speech perception [15]. Our results are remarkably in line with a Bayesian model of perception where stronger priors engender a bias towards hallucinations and foster perceptual confidence in light of ambiguous sensory input.

Materials and methods

Participants

Lab experiment. The final sample of the lab experiment comprised *N = 37* participants of which eleven were female; age ranged from 18 to 30 [mean 21.97] years. Exclusion criteria were self-reported hearing loss; neurological or psychiatric disorders; or the regular consumption of drugs, particularly amphetamines, cannabis, or similar psychoactive substances. Originally, *N* = 42 volunteers had been recruited, but data from five of these had to be excluded due to software problems during testing.

Online study. The final sample of the online study comprised *N* = 131 participants, of which 94 participants were female; age ranged from 19–61 [mean 27.2] years. A total of *N* = 353 volunteers had been recruited through dissemination of the study link in social media, university e-mail lists, and psychology student councils of different German Universities. *N* = 131 complete datasets were obtained. Self-exclusion was based on the same criteria as in the lab experiment.

All participants gave informed consent. Psychology students at the University of Lübeck were offered to obtain course credit. All procedures were approved by the local ethics committee of the University of Lübeck.

Experimental Procedures

In the lab, participants first performed the psychoacoustic experiment, and then completed two questionnaires, namely the Schizotypal Personality Questionnaire (SPQ) and the Launay-Slade hallucination scale (LSHS). The rationale was to perform the relatively long psychoacoustic experiment first in order to not unnecessarily tire participants with the questionnaires before. Psychoacoustic testing was performed on a PC. Stimuli were delivered through Sennheiser HD 280 Pro headphones. Presentation level was kept constant at a comfortable and clearly audible level. In total, the duration of the lab experiment was approximately 60–75 minutes.

In the online experiment, the order of questionnaires and psychoacoustics was reversed, that is, participants first filled out the two questionnaires and afterwards performed the psychoacoustic experiment. The drop-out rate is expected to be higher in online than in lab experiments. Thus, the rationale for reversing the order was that participants dropping out after the questionnaires we would at least deliver their SPQ and LSHS data. Also, the duration of the psychoacoustic online experiment was reduced to 20 % of duration of the lab experiment (speechiness kernels were shown to be highly similar with 100 and 540 trials, see supplementary Fig. S1), leaving the psychoacoustic testing shorter and less tiring. The order of the two questionnaires was randomized across participants.

For the psychoacoustic task, participants were instructed to use headphones. Prior to the psychoacoustic task, participants could adjust the presentation intensity to a comfortable level using an exemplar sound texture and were instructed to keep the presentation level constant during the experiment. The total duration of the experiment amounted to approximately 20 - 25 min. Participants were debriefed after the experiment.

Questionnaires

Schizotypy was assessed using the German adaptation of the schizotypal personality questionnaire (SPQ; Raine 1991, Klein, Andresen, Jahn, 1997). The SPQ comprises nine subscales based on *DSM-III-R* criteria for diagnosis of schizotypal personality disorder, namely: ideas of reference; excessive social anxiety; odd beliefs or magical thinking; unusual perceptual experiences; odd or eccentric behavior; no close friends; odd speech; constricted affect; and suspiciousness. In the current study, we were particularly interested in the subscale ‘unusual perceptual experiences’ (SPQ-UP) to measure predisposition to hallucinations. In total, the SPQ includes 74 items of which the responses (true/false) are summed up to derive a total score, amounting to a maximal total score of 74. The SPQ-UP subscale has a maximum total score of nine.

Predisposition for hallucinations was assessed using the German version of the Launay-Slade Hallucination Scale (LSHS; Launay & Slade 1981; Lincoln et al., 2009). The questionnaire comprises twelve items which are assessed on a five-point Likert scale (0 – 4). The LSHS score is derived as the sum of all items and can thus maximally reach 48.

Psychoacoustic testing

Stimuli. Stimuli were resynthesized natural sounds (“sound textures” [18]) presented in white noise at an SNR of 3 dB. Textures were synthesized from the spectro-temporal modulation content of a large set of real-life sounds (*n* = 192), including speech, voice, animal vocalizations as well as nature and tool (instrument) sounds we had used in a previous study [13] (see Fig. 1a,b). Texture synthesis parameters were as follows: frequency range 0.02–10 kHz, number of frequency bands = 30, sampling rate = 20 kHz, temporal modulation range 0.5–200 Hz, sampling rate = 400 Hz; maximum number of iterations = 60. Textures had a length of 1 s and a sampling rate of 20 kHz. Out of the total of 192 textures, we selected the 108 “most speech-like” textures for the final experiment: “speech-like” textures were defined as those textures whose spectral centroid diverted less than 3 standard deviations from the mean spectral centroid of those textures that had originally been synthesized from speech stimuli.

Task. Two sounds were randomly paired on each trial and presented with an inter-stimulus-interval of 500 ms. In a two-alternative-forced-choice (2-AFC) task, participants were asked to decide “which exemplar sounded more like speech” (Fig. 1c). Participants rated their confidence from 1–3 (unconfident to confident).

The lab experiment was performed in Matlab 2018b. The order of exemplars was randomized across participants. In total, each texture exemplar was presented five times. The first ten trials were training trials. Twenty relatively unambiguous catch trials were distributed evenly across the experiment. In both catch and training trials, a speech texture (that is, a texture of which the spectral centroid diverted < 0.6 from the mean spectral centroid of speech textures) was paired with a texture from the other categories. In total, the lab experiment comprised 560 trials, including 20 catch trials. Catch and training trials were excluded from the subsequent reverse correlation analyses, such that speechiness kernels were estimated based on 540 trials.

The online experiment was identical to the lab experiment with the following exceptions: the online experiment was performed in Labvanced (Scicovery GmbH, Osnabrück, Germany). The order of exemplars was fixed across participants due to programming constraints in Labvanced. To keep the experiment short (approximately 10 min), each texture exemplar was presented only once. The first four trials were training trials. Five relatively unambiguous catch trials were distributed evenly across the experiment. In total, the online experiment comprised 117 trials out of which of 5 were catch trials. Catch and training trials were excluded from the subsequent reverse correlation analyses, such that speechiness kernels were estimated based on 108 trials. Note that kernels estimated based on 100 trials are highly similar to the ones based on 540 trials, but for details on kernel stability see supplementary Fig. S1.

Analyses

Sound decomposition. We analyzed how participants' judgements varied as a function of the spectro-temporal modulation content of the stimuli. The modulation content of the stimuli was obtained by filtering the sounds with a model of auditory processing [12] using the "NSL Tools" package (available at http://www.isr.umd.edu/Labs/NSL/Software.htm) and customized Matlab code (The MathWorks Inc., Matlab 2014b/2018a). First, spectrograms for all sounds were obtained using a bank of 128 overlapping bandpass filters with equal width (Q_{10dB} = 3), spaced along a logarithmic frequency axis over a range of f = 116–2872 Hz (hair cell stage). A midbrain stage modelled the enhancement of frequency selectivity as a first-order derivative with respect to the frequency axis, followed by a half-wave rectification and a short-term temporal integration (time constant τ = 8 ms).

Then, the auditory spectrogram was analyzed by the cortical stage, where the modulation content of the auditory spectrogram was computed through a bank of 2-dimensional filters selective for a combination of spectral and temporal modulations. The filter bank performs a complex wavelet decomposition of the auditory spectrogram. The magnitude of such decomposition yields a phase-invariant measure of modulation content. The modulation selective filters have joint selectivity for spectral and temporal modulations, and are directional, i.e. they respond either to upward or downward frequency sweeps.

Modulation filters. Filters were tuned to spectral modulation frequencies of Ω = [0.5, 0.76, 1.15, 1.74, 2.64, 4] cyc/oct, temporal modulation frequencies of ω = [1, 2, 4, 8, 16, 32] Hz, and centre frequencies of f = [116, 183, 290, 459, 726, 1148, 1816, 2872] Hz. The filter bank output was computed at each frequency along the tonotopic axis and then averaged over time. The time-averaged output of the filter bank was averaged across the upward and downward filter directions. This resulted in a representation with 6 spectral modulation frequencies, 6 temporal modulation frequencies, and 8 tonotopic frequencies, amounting to 288 acoustic features in total. The rationale for this choice of values was to use a decomposition roughly covering the temporal and spectral modulations present in speech (for spectro-temporal modulation content of all sound categories see Fig. 1a, for modulation content of speech see Fig. 1b).

Psychophysical reverse correlation. To obtain an internal template of speech, we used the reverse correlation technique [37,38]. Sounds were sorted into speech and nonspeech stimuli based on the participants' responses. Note that due to the 2-AFC task, the two choices are symmetric, that is, on each trial one stimulus was attributed to the speech and hence the other one to the nonspeech category. Therefore, to obtain individual speechiness kernels we subtracted the two templates for speech and nonspeech:

$$K(f)= E[s(f)|\text{speech}] - E[s(f)|\text{nonspeech}] \qquad (1)$$

where $E[s(f)|\text{speech}]$ indicates the trial average of the stimulus at feature f conditional on choice "speech", $s(f)$ is the stimulus at feature f, and $K(f)$ is the magnitude of the psychophysical kernel at feature f.

Kernel stability. To obtain an estimate of how many trials are necessary to obtain a stable estimate of individual speechiness kernels, we assessed kernel stability in the lab-experiment data. To this end, we iteratively calculated the Pearson's correlation between the kernel based on a subset of n trials and the final kernel based on all 540 trials [39] for all participants.

Principal component analysis (PCA). To reduce dimensionality of the speechiness kernel (comprising n = 288 acoustic features), we computed the matrix singular value decomposition on the set of all lab and online speechiness kernels (all scaled between 0 and 1) jointly in Matlab 2018a and jamovi 1.0.7.0. This yielded k =

288 components. We decided on the number of components to retain based on a scree plot [40] and parallel analysis [41]. We retained only those components where the eigenvalues associated with the raw data exceed the eigenvalues of surrogate data, leaving us with the first six components (see scree plot Fig. 4b). The individual speechiness kernels were then projected through (i.e., multiplied with) those components. The individual component loadings were scaled by the singular values, yielding individual component scores. Lastly, the component scores were correlated using Pearson's *r* with the LSHS and SPQ-UP subscale, respectively. The two scales (LSHS and SPQ-UP subscale) were chosen as they measure psychotic-like experiences, specifically predisposition for unusual perceptual experiences (e.g., hallucinations).

Sound discriminability analysis. To estimate the discriminability of two textures presented on a trial, we first calculated trial-wise sound pair distance as the Euclidian distance between each of the 288 features. To investigate whether individual speechiness kernels influenced the discriminability of sound pairs presented on each trial, sounds in the modulation space (comprising 288 features) were projected through (multiplied with) individual speechiness kernels. Then, for each sound separately, the features were normalized between their minimum and maximum (effectively scaling them between 0 and 1). Sound pair distance was compared before and after projection through speechiness kernels.

Generalized linear mixed-effects models. To evaluate the relation of trial-wise confidence judgements with sound-pair discriminability (see above) and hallucination proneness, we used a Cumulative Link Mixed Models, that is, a hierarchical generalized linear model with a cumulative link function (i.e., an ordinal regression model). Ordinal linear mixed-effects regression was fitted using the Bayesian estimation package *brms* in RStudio 1.3.959 (cumulative-probit link function). We regressed trial-by-trial confidence judgements (on a Likert scale from 1 ["not confident"] to 3 ["confident"]) against the *z*-scored predictors trial-wise sound pair distance, LSHS score, global SPQ score, as well as experiment type (online, lab) as covariate of no interest with subject-specific random intercepts.

Acknowledgements

This research was funded by an ERC consolidator Grant (ERC-CoG-2014-646696 "AUDADAPT" to J.O.). We thank Felix Greuling for the implementation of the online study and Tabea Landschoff, Franziska Mauz, Hannah Kleinhaus for data acquisition.

References

1 de Leede-Smith, S. & Barkus, E. A comprehensive review of auditory verbal hallucinations: lifetime prevalence, correlates and mechanisms in healthy and clinical individuals. *Frontiers in human neuroscience* **7**, 367, doi:10.3389/fnhum.2013.00367 (2013).

2 Sommer, I. E. *et al.* Healthy individuals with auditory verbal hallucinations; who are they? Psychiatric assessments of a selected sample of 103 subjects. *Schizophrenia bulletin* **36**, 633-641, doi:10.1093/schbul/sbn130 (2010).

3 Corlett, P. R. *et al.* Hallucinations and Strong Priors. *Trends Cogn Sci* **23**, 114-127, doi:10.1016/j.tics.2018.12.001 (2019).

4 Teufel, C. *et al.* Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13401-13406, doi:10.1073/pnas.1503916112 (2015).

5 Powers, A. R., Mathys, C. & Corlett, P. R. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* **357**, 596-600, doi:10.1126/science.aan3458 (2017).

6 Alderson-Day, B. *et al.* Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. *Brain : a journal of neurology* **140**, 2475-2489, doi:10.1093/brain/awx206 (2017).

7 Horga, G., Schatz, K. C., Abi-Dargham, A. & Peterson, B. S. Deficits in predictive coding underlie hallucinations in schizophrenia. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **34**, 8072-8082, doi:10.1523/JNEUROSCI.0200-14.2014 (2014).

8 Dierks, T. *et al.* Activation of Heschl's gyrus during auditory hallucinations. *Neuron* **22**, 615-621 (1999).

9 Allen, P., Laroi, F., McGuire, P. K. & Aleman, A. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neuroscience and biobehavioral reviews* **32**, 175-191, doi:10.1016/j.neubiorev.2007.07.012 (2008).

10 Formisano, E. *et al.* Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* **40**, 859-869 (2003).

11 Dau, T., Kollmeier, B. & Kohlrausch, A. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America* **102**, 2892-2905 (1997).

12 Chi, T., Ru, P. & Shamma, S. A. Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America* **118**, 887-906 (2005).

13 Erb, J. *et al.* Homology and Specificity of Natural Sound-Encoding in Human and Monkey Auditory Cortex. *Cerebral cortex* **29**, 3636-3650, doi:10.1093/cercor/bhy243 (2019).

14 Santoro, R. *et al.* Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 4799-4804, doi:10.1073/pnas.1617622114 (2017).

15 Uhlhaas, P. J. & Mishara, A. L. Perceptual anomalies in schizophrenia: integrating phenomenology and cognitive neuroscience. *Schizophrenia bulletin* **33**, 142-156, doi:10.1093/schbul/sbl047 (2007).

16 Bizley, J. K. & Cohen, Y. E. The what, where and how of auditory-object perception. *Nature reviews. Neuroscience* **14**, 693-707, doi:10.1038/nrn3565 (2013).

17 Elliott, T. M. & Theunissen, F. E. The modulation transfer function for speech intelligibility. *PLoS computational biology* **5**, e1000302, doi:10.1371/journal.pcbi.1000302 (2009).

18 McDermott, J. H. & Simoncelli, E. P. Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron* **71**, 926-940, doi:10.1016/j.neuron.2011.06.032 (2011).

19 Murray, R. F. Classification images: A review. *Journal of vision* **11**, doi:10.1167/11.5.2 (2011).

20 Ahumada, A. & Lovell, J. Stimulus features in signal detection. *The Journal of the Acoustical Society of America* **49**, 1751–1756 (1971).

21 Raine, A. *et al.* Cognitive-perceptual, interpersonal, and disorganized features of schizotypal personality. *Schizophrenia bulletin* **20**, 191-201, doi:10.1093/schbul/20.1.191 (1994).

22 Klein, C., Andresen, B. & Jahn, T. Erfassung der schizotypen Persönlichkeit nach DSM-III-R: Psychometrische Eigenschaften einer autorisierten deutschsprachigen Übersetzung des „Schizotypal Personality Questionnaire“ (SPQ) von Raine. *Diagnostica* **43**, 347–369, doi:https://doi.org/10.1037/t10727-000 (1997).

23 Launay, G. & Slade, P. The measurement of hallucinatory predisposition in male and female prisoners. . *Personality and Individual Differences* **2**, 221–234, doi:https://doi.org/10.1016/0191-8869(81)90027-1 (1981).

24 Lincoln, T. M., Keller, E. & Rief, W. Die Erfassung von Wahn und Halluzinationen in der Normalbevölkerung: Deutsche Adaptationen des Peters et al. Delusions Inventory (PDI) und der Launay Slade Hallucination Scale (LSHS-R). *Diagnostica* **55**, 29–40, doi:https://doi.org/10.1026/0012-1924.55.1.29 (2009).

25 Laroi, F. How do auditory verbal hallucinations in patients differ from those in non-patients? *Frontiers in human neuroscience* **6**, 25, doi:10.3389/fnhum.2012.00025 (2012).

26 Barkus, E. *et al.* Auditory false perceptions are mediated by psychosis risk factors. *Cognitive neuropsychiatry* **16**, 289-302, doi:10.1080/13546805.2010.530472 (2011).

27 Parola, A., Simonsen, A., Bliksted, V. & Fusaroli, R. Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis. *Schizophrenia research* **216**, 24-40, doi:10.1016/j.schres.2019.11.031 (2020).

28 Neri, P. & Levi, D. M. Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision research* **46**, 2465-2474, doi:10.1016/j.visres.2006.02.002 (2006).

29 Horga, G. & Abi-Dargham, A. An integrative framework for perceptual disturbances in psychosis. *Nature reviews. Neuroscience* **20**, 763-778, doi:10.1038/s41583-019-0234-1 (2019).

30 Teufel, C. & Fletcher, P. C. Forms of prediction in the nervous system. *Nature reviews. Neuroscience* **21**, 231-242, doi:10.1038/s41583-020-0275-5 (2020).

31 Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D. & Friston, K. J. The computational anatomy of psychosis. *Frontiers in psychiatry* **4**, 47, doi:10.3389/fpsyt.2013.00047 (2013).

32 Adams, R. A., Huys, Q. J. & Roiser, J. P. Computational Psychiatry: towards a mathematically informed understanding of mental illness. *Journal of neurology, neurosurgery, and psychiatry* **87**, 53-63, doi:10.1136/jnnp-2015-310737 (2016).

33 Mathys, C. D. *et al.* Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in human neuroscience* **8**, 825, doi:10.3389/fnhum.2014.00825 (2014).

34 Sedley, W., Friston, K. J., Gander, P. E., Kumar, S. & Griffiths, T. D. An Integrative Tinnitus Model Based on Sensory Precision. *Trends Neurosci* **39**, 799-812, doi:10.1016/j.tins.2016.10.004 (2016).

35 Baumann, S. *et al.* The topography of frequency and time representation in primate auditory cortices. *eLife* **4**, doi:10.7554/eLife.03256 (2015).

36 Erb, J., Schmitt, L. M. & Obleser, J. Temporal selectivity declines in the aging human auditory cortex. *BioRxiv* doi:doi.org/10.1101/2020.01.24.919126. (2020).

37 Neri, P. & Heeger, D. J. Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature neuroscience* **5**, 812-816, doi:10.1038/nn886 (2002).

38 de Boer, R. & Kuyper, P. Triggered correlation. *IEEE transactions on bio-medical engineering* **15**, 169-179, doi:10.1109/tbme.1968.4502561 (1968).

39 Burred, J. J., Ponsot, E., Goupil, L., Liuni, M. & Aucouturier, J. J. CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. *PloS one* **14**, e0205943, doi:10.1371/journal.pone.0205943 (2019).

40 Cattell, R. B. The Scree Test For The Number Of Factors. *Multivariate behavioral research* **1**, 245-276, doi:10.1207/s15327906mbr0102_10 (1966).

41 Horn, J. L. A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika* **30**, 179-185, doi:10.1007/BF02289447 (1965).