Functional significance of spectrotemporal response functions obtained using magnetoencephalography

Francisco Cervantes Constantino[1], Marisel Villafañe-Delgado[2], Elizabeth Camenga[2], Katya Dombrowski[2], Benjamin Walsh[2], Jonathan Z. Simon[2,3,4] *

[1]Program in Neuroscience and Cognitive Science, University of Maryland, College Park, Maryland, United States of America

[2]Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland, United States of America

[3]Department of Biology, University of Maryland, College Park, Maryland, United States of America

[4]Institute for Systems Research, University of Maryland, College Park, Maryland, United States of America

*Corresponding author: jzsimon@umd.edu

Short Title: Non-invasive Human STRF models

Abstract

The spectrotemporal response function (STRF) model of neural encoding quantitatively associates dynamic auditory neural (output) responses to a spectrogram-like representation of a dynamic (input) stimulus. STRFs were experimentally obtained via whole-head human cortical responses to dynamic auditory stimuli using magnetoencephalography (MEG). The stimuli employed consisted of unpredictable pure tones presented at a range of rates. The predictive power of the estimated STRFs was found to be comparable to those obtained from the cortical single and multiunit activity

literature. The STRFs were also qualitatively consistent with those obtained from electrophysiological studies in animal models; in particular their local-field-potential-generated spectral distributions and multiunit-activity-generated temporal distributions. Comparison of these MEG STRFs with others obtained using natural speech and music stimuli reveal a general structure consistent with common baseline auditory processing, including evidence for a transition in low-level neural representations of natural speech by 100 ms, when an appropriately chosen stimulus representation was used. It is also demonstrated that MEG-based STRFs contain information similar to that obtained using classic auditory evoked potential based approaches, but with extended applications to long-duration, non-repeated stimuli.

Author summary

The spectrotemporal response function (STRF) model of linking dynamic acoustic stimuli to dynamic neural responses is applied to whole-head non-invasive magnetoencephalography (MEG) recordings of the human auditory cortex. MEG STRFs were consistent predictors of neural activity, quantitatively and qualitatively, by comparison to those obtained from animal models using local field potential or multiunit activity as neural responses. Comparison of STRFs using stimuli as diverse as tone clouds, natural speech, and music revealed a common structure consistent with shared baseline auditory processing, when an appropriately chosen stimulus representation was used.

Introduction

Empirically measured sensory receptive fields and response functions offer analytical characterizations of computations attainable by the auditory system[1–3]. Applied linear systems methods such as the spectrotemporal response function (STRF)[4–6] have similarly led to informative computational characterizations of central auditory neural function with respect to sound encoding and perception[7]. The STRF can be viewed as a representation of the approximate neural response to changing auditory input in time or frequency; any particular functional description will vary according to the location and role of the neurons. Different stimulus classes (e.g. artificially generated sounds vs. natural sounds) may produce related, but dissimilar STRFs from the same neural unit, speaking to fundamental processing differences (and similarities) of auditory encoding[8–10]. An emerging view in electrophysiology is that the STRF may represent a snapshot of the entire network converging onto that neuron (or group of neurons)[10], incorporating this population's activity in its neural representation of the spectrotemporal features of the stimulus[7]. As seen here, STRFs also have a role in investigations of ensemble auditory coding, using neural recordings obtained from magnetoencephalography (MEG) or electroencephalography (EEG).

STRFs directly characterize the relationship between a sound stimulus and the accompanying neural response. For neural ensembles, rather than individual neurons, many individual linear components may be jointly pooled, perhaps even superadditively (depending on the underlying neuroanatomy and neurophysiology of the signal source).

Also, as in the case of a single-neuron-based STRF, it may be methodologically simpler to use controlled stimuli rather than natural sounds[11–13]. It remains to be determined how the spectrotemporal features of ensemble-based STRFs correspond to the time-varying evoked-related-potential responses (and other standard MEG/EEG measures) as a function of frequency, and also to what extent the STRF encoding model can provide analogous additional information besides predictive power. Furthermore, the STRF estimate of a stimulus-response relationship may depend on the particular representation chosen for that stimulus; in particular, it remains unknown which specific stimulus representations are optimal for the purpose of matching STRF features to neural function[14], and whether any such choices can address the key question of how to generalize across stimulus classes, from artificial to natural stimuli. Finally, it is important to discuss overlap between these non-invasively obtained STRFs and those available from local field potential (LFP) data or from other invasive recordings.

In order to address these questions, evoked cortical activity recordings from healthy listeners were obtained with MEG during active listening of pseudo-random multi-tone patterns[12,15] presented at a variety of rates. STRFs were obtained per subject and condition, in order to assess the extent to which the MEG responses were linearly explainable by a sparse representation of the stimulus sound pattern, and whether rate-related changes are consistent with those found using invasive electrophysiological techniques. Peak components in STRFs and temporal response functions (TRFs) were identified and their latencies compared to those obtained with standard tone-based averaging. Alternative representations of the stimulus, including the auditory spectrogram, were used for reverse correlation in order to constrain the space of stimulus

representations given the properties of the MEG cortical signal. Finally, these functionally informative STRFs were compared to those from datasets from studies using natural speech[16] and music processing[17]. This allowed an investigation of ensemble-dependent issues arising from STRF comparisons when using artificial vs. natural stimuli[9].

MEG-based STRFs are shown to functionally explain considerable amounts of response variability while revealing a parsimonious mapping of response features seen in classic averaging methods to those obtained from dynamic stimuli timeseries. Quantitatively, the MEG-based STRFs account for similar levels of predictive power to single and multiunit responses in auditory cortex[11]. Qualitatively, the mappings show reasonable correspondence with those from local field potential activity in animal models[18,19] and manifest similar stimulus dependencies (e.g., density[12]). We find that similar STRF structure is seen across responses to stimuli as diverse as natural speech and music, demonstrating convergence across stimulus classes. This last result, however, depends on the use of a specific (sparse) representation of acoustic stimuli, the nature of which provides additional knowledge regarding the role of spectrotemporal modulations on predictive frameworks of auditory cortical representations over a wide range of dynamic sound classes.

Results

MEG cortical responses predictable from the STRF linear model

Potential successes of the STRF as a linear model to predict MEG responses from acoustic stimuli are evaluated by comparing the actual vs. predicted responses which, unlike spike-generated STRFs, are continuous waveforms (Fig 1A). Model predictions are obtained by linearly convolving the corresponding STRF with the stimulus representation, using cross-validation (arbitrary separation of training data from testing data) to prevent overfitting, which makes this a conservative estimate due to noise present in the testing data[11,13]. If instead only the training data is used, i.e., fitting to the same data as is tested, STRF estimates provide a stringent upper limit as to how good any linear prediction can be. STRFs estimated using cross-validation predict the large negative deflections (Fig 1a, red) that follow tone onsets (~100 ms post impulse) well, but unlike those from training (Fig 1a, blue), are less accurate for positive excursions (both data sets summarized in Fig 1b). The ability of the STRF model to predict the encoding relationship between sound patterns and cortical responses can be measured as the fraction of response variability explainable by the linear model, estimated on an individual condition and subject basis, once intrinsic response variability (unrelated to the stimulus) has been removed[11,13]. MEG STRF predictions were found to range as high as 34% of variance explained across participants and rates, using cross-validated data. When the fraction of variance explainable by the model was compared with normalized noise power (or inverse SNR), the explainable fraction in the theoretical noiseless limit was estimated to be 23.0 ± 2.0% (mean ± st. dev.; CI: 19.0–26.9%) as part of a significant linear regression relationship (F=45.9; p=2.7x10^{-9}; R^2=0.386), with an upper limit of 71% as provided by training-data only results (Fig 1c).

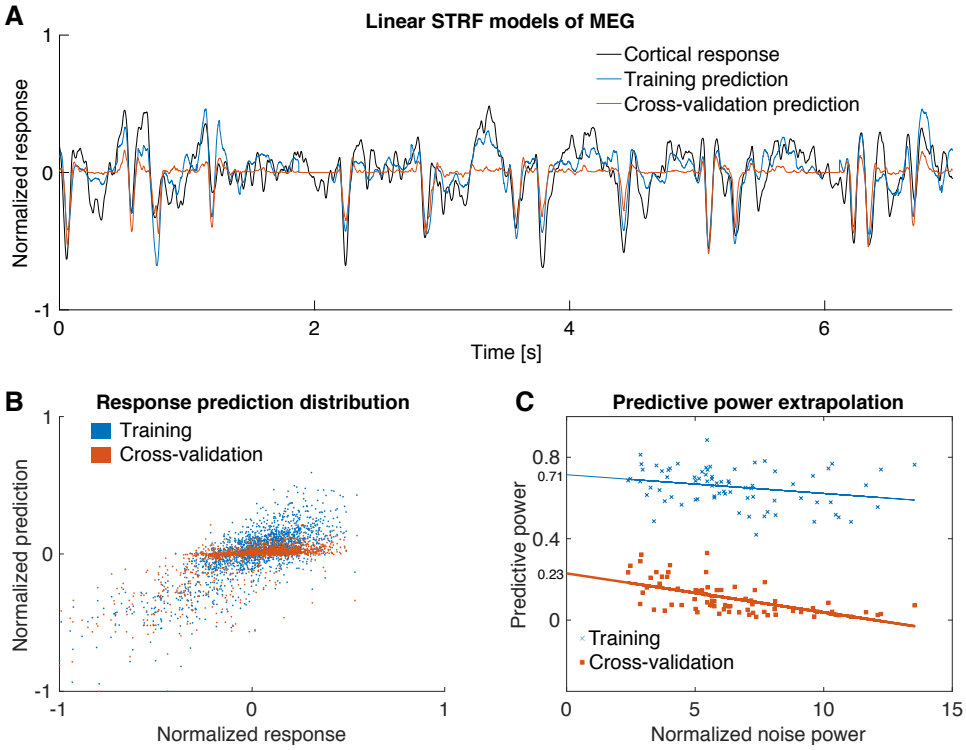

Fig 1. Spectrotemporal encoding models of MEG signals from human auditory cortex. a) A 7 s sample recording of an MEG response to a sparse multitone pattern (2 tones/s), with STRF-based predictions. b) STRFs were optimized by iteratively minimizing prediction error on the entire dataset, referred to as *training* (blue, r=0.74), or alternatively on their ability to generalize (*cross-validate*) over testing datasets (red, r=0.62). c) The predictive power of the STRF models is shown by linear regression of individual STRFs across participants and conditions on their corrected normalized noise power (i.e., inverse signal-to-noise ratio, an indicator of trial by trial reliability, see *Methods*). Extrapolation of performance to zero noise power gives a noise-corrected expected performance for both the conservative cross-validation-based estimates and the fundamental-upper-limit training estimates.

Fraction of response explained by STRF features consistent with standard evoked potentials

STRFs based on MEG responses display consistent spectrotemporal structure in the form of positive-negative-positive complex deflections (Fig 2a) coinciding with typical auditory cortical latencies (e.g. those of the P1-N1-P2 complex in averaged EEG responses to isolated tones). In particular, the multitone STRFs demonstrate strong negative responses at ~100 ms post impulse onset ($STRF_{100}$). The specific $STRF_{100}$ latency depends on stimulus frequency, varying ~20 ms over the frequency range 180-700 Hz; at higher frequencies the latency is approximately constant ($STRF_{100}$ latencies for 2 tones/s shown in Fig 2b, black). $STRF_{100}$ latencies were found to follow standard tone-evoked M100 latencies[20–25] obtained under various conditions (Fig 2b; also Table 1). The correspondence suggests a quantitative link between the $STRF_{100}$ and M100, and therefore between STRF-based techniques and ordinary auditory evoked cortical potentials. Analyzing the same experimental data using standard evoked response analysis instead (epoching and averaging over responses to all tones in the sparsest multitone pattern) demonstrates strong temporal correspondence at the group level (S1 Fig).

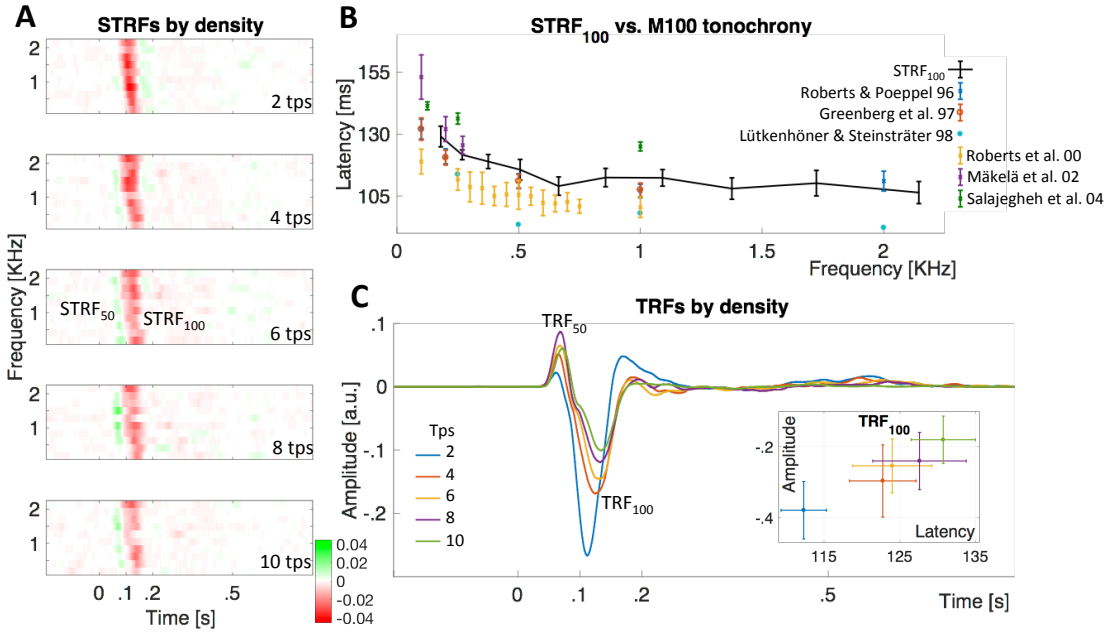

Fig 2. Consistency between response function model predictive model features and evoked potentials. a) Grand average spectrotemporal response functions based on multitone stimuli demonstrate a positive-negative-positive structured sequence between 50 and 200 ms following tone onset; tone cloud density introduces qualitative changes in relative amplitude and delays: at increased rates an early positive component (50-100 ms; $STRF_{50}$) emerges, while the medium latency negative component (100-150 ms; $STRF_{100}$) attenuates, and a late positive component (150-200 ms) present only in the sparsest conditions disappears. b) $STRF_{100}$ components are delayed by over 20 ms as tone carrier frequency decreases from 2 to 0.2 KHz, in a manner consistent with those of evoked potentials in single tone presentations[20,21,23–25] (Table 1), indicating a correspondence between impulse response functions obtained through reverse correlation

and averaged evoked potentials. A common latency decrease across studies and conditions is observed for carrier frequencies in the 180-700 Hz range. c) Temporal response functions, obtained by reverse correlation with the stimulus envelope collapsed across frequencies, show features similar to the P1-N1-P2 complex commonly found in EEG evoked potentials[26]. Higher tone presentation rates result in the emergence of the TRF_{50} and in decreased amplitude and increased latency of the TRF_{100} (inset), as well as the attenuation of a later-latency positive deflection. Error bars are 1 standard error of the mean.

	# of subjects [mean age]	**Sound delivery [Sensor location]**	**Tone duration (ms)**	**Presentation rate [tones/s]**	**Peak finding method**
Roberts & Poeppel, 1996[20]**; Greenberg et al., 1997**[21]	5 [24-33 y]	Monaural [Contralateral, Left]	400	0.7 - 1.3	Equivalent dipole, Maximum RMS
Lütkenhöner & Steinsträter, 1998[22]	1 [28 y]	Monaural [Contralateral, Left]	520	~0.4	Maximum RMS
Roberts et al., 2000[23]	8 [-]	- [Both hemispheres]	-	-	-
Mäkelä et al., 2002[24]	11 [32 y]	Binaural [Both hemispheres]	200	1	Optimal sensor pair
Salajegheh et al., 2004[25]	11 [45.8 y]	Binaural [Both hemispheres]	400	0.8 - 1.3	Maximum RMS from optimal 12 sensors

Table 1. Comparison of studies reporting M100 absolute latency values in response to pure tones, with participants, recording mode, stimulus details, and M100 peak determination method where available.

To further investigate the correspondence between STRFs and evoked potentials (specifically the effects of tone density), reverse correlation was performed with respect to frequency-collapsed representations of the stimulus, generating the frequency-independent Temporal Response Function (TRF, Fig 2c). The ~100 ms latency negative peak (TRF_{100}) amplitude decreased with increasing tone-density by ~60% across modulation rate range studied, while latency increased 20% (see inset). In contrast, the ~50 ms latency positive deflections (TRF_{50}) had the smallest amplitude for the sparsest multitone condition. Thus sources with ~50 ms latency generate a strong increase in cortical activity with a transition from scattered to continuous pure tones, while sources with ~100 ms latency decrease in strength as they are delayed. Cortical activity in sources with 150 ms latency may also be active, provided the inter-tone interval is long.

STRF most informative for onset-based representations of multitone stimulus

Methodologically, the acoustic representation of the stimulus used to generate the STRF may employ any number of available time-frequency representations of the sound, including the widely-used spectrogram[11,27–29]. One reason to consider alternatives to the spectrogram is to compare STRF features with evoked response features, since an evoked response to tones is calculated not with respect to the spectrotemporal duration of the tones but only to their onsets. Thus analyses also included binary and sparse

representations of the stimulus: single tones were modeled as trigger-like impulses timed with tone onset and organized by frequency. Indeed, stimulus features that are known to be encoded by auditory cortex include onsets, offsets, and stimulus duration (in the form of sustained responses)[30–33]. Since the MEG signal is aggregated across synchronized individual neurons[34], evidence for those same encodings requires investigation. Reverse correlation techniques are well suited for this larger-scale analysis because it explores the outcome of alternative stimulus representations that emphasize such features. The stimulus representations tested here (cf. Fig 3 insets) were (i) the ideal *trigger* representation, (ii) the ideal *edge* representation (both onset and offset triggers), (iii) the ideal stimulus first-order *derivative* (onset and negatively-signed-offset triggers), which can itself be used to generate the trigger representation if followed by half-wave rectification, (iv) the ideal stimulus *pulse* envelope, which has constant value from onset to offset (and which can itself be used to generate the previous representation if followed by differentiation), (v) the actual acoustic stimulus passed through a filterbank with identical center frequencies as the tone, whose *envelope* is then extracted (see *Methods*), and (vi) a generalized *envelope onset* representation obtained via half-wave rectification of the previously defined filterbank envelope output. Only the last two can be applied to natural (non-discrete) stimuli, and so are especially important in later sections.

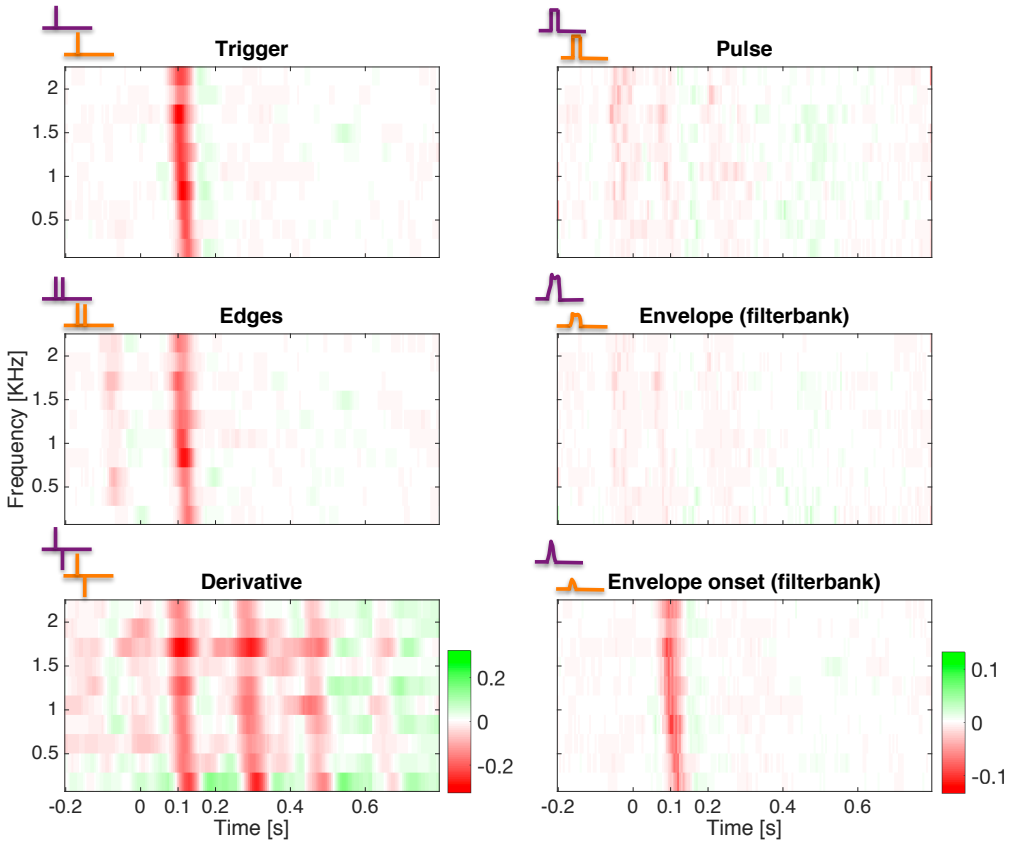

Fig 3. STRFs generated using different stimulus representations achieve different levels of functionality. STRFs generated from multitone patterns are functionally informative (e.g., comparable to evoked potential analysis) when each individual tone is discretely represented by its onset (top left) but not when represented instead by the timing of its temporal edges (middle left), sign (bottom left), or a discrete representation of the entire pulse duration (top right). Related to the spectrogram, the representation based on passing the acoustic signal through a series of filterbanks, then extracting envelopes per band (middle right) yields only barely discernible results. Extracting onset timing information from the filterbank, in contrast, was quite functionally informative (half-wave rectification of the first derivative of the filterbank output; bottom left). Critically, filterbank-based methods do not require *a priori* definitions of temporal edges

and can be used for arbitrary stimuli. Color scales as in bottom right inset, except for Derivative STRF.

Grand average STRFs in Fig 3 demonstrate that among such representations, only those expressing tone onset events explicitly yield components comparable to those of evoked potential analysis (first and last STRFs of Fig 3); STRFs from the alternatives introduced ringing and/or pre-causal artifacts. As with the original onset-based trigger representation, reverse correlation with both temporal edges predicted activity from the first edge in accordance with the latency by frequency dependence, but also produced a pre-causal mirror component, in advance of the original and at the tone duration distance. This pattern suggests that tone offset was not explicitly encoded here. This interpretation is supported by analysis of STRFs generated by the derivative representation, which correspondingly flips the sign of the same pre-causal mirror component, but is additionally contaminated via constructive interference by a series of artefactual ringing cycles. The pulse representation, which can be viewed as an idealized envelope, produces STRFs that are essentially featureless (or at best, whose features are barely discernible above the noise floor). This result is unexpected since typical auditory reverse correlation studies use a duration-based stimulus envelope representation[35,36] and the temporal envelope is often hypothesized to be the response-driving feature. Similarly, the acoustic envelope representation (using a filterbank model; see *Methods*) also produced featureless STRFs. An attempt to re-create the onset representation (i.e. half-wave rectification of the acoustic envelope representation derivative), did however generate STRFs with features comparable to evoked potential analysis, and enables the extraction

of onset-like information in general from diverse complex natural stimuli. Because of the remarkable agreement between the idealized and the acoustic onset models, interpretations based on evoked potentials may extend to reverse correlation analysis applied to other stimulus classes where definitions of onsets would be *a priori* unknown or not controlled for, such as natural sounds.

Convergent STRF models across artificial and natural stimuli

Because of its potential to reveal hierarchical processing mechanisms, a major goal in auditory reverse correlation has been to examine the encoding relationship for critical natural stimuli including speech and other communication sounds. To this end, datasets from two previously unpublished studies on speech and music processing were submitted to the same analysis methods as the multitone pattern (Fig 4a), with stimuli represented by their envelope onsets. As with the onset-based representation of the multitone patterns, STRFs for speech and music exhibited qualitatively similar structures, with distinctive biphasic components near 50 and 100 ms post rising transient impulses (onsets) along the same investigated spectral region. Inspection of the stimuli under either envelope or envelope onset representations suggests that the latter procedure effectively increases similarity in the underlying distribution across stimulus classes (S2a Fig). The frequency dependence of relative peak delays was also maintained for these stimulus classes (S2b Fig) but with class-dependent timing differences, suggesting a common fundamental

mode of spectrotemporal cortical processing up to ~200 ms and after which notable processing differences appear according to the stimulus class. While neural data from all studies were obtained from different subject groups, one subject did participate in those two studies and in a modified pilot version of this experiment (2.4 tones per second presentation rate); these data are presented in Fig 4b, again showing strong qualitative similarity both to group data and class-dependent timing differences. This subject's topographic magnetic field maps associated with the neuromagnetic signals derived in each of the three studies are displayed in Fig 4c; mapping each STRF to overlapping spatial distributions is consistent with source activity at the superior aspect of the temporal lobes.

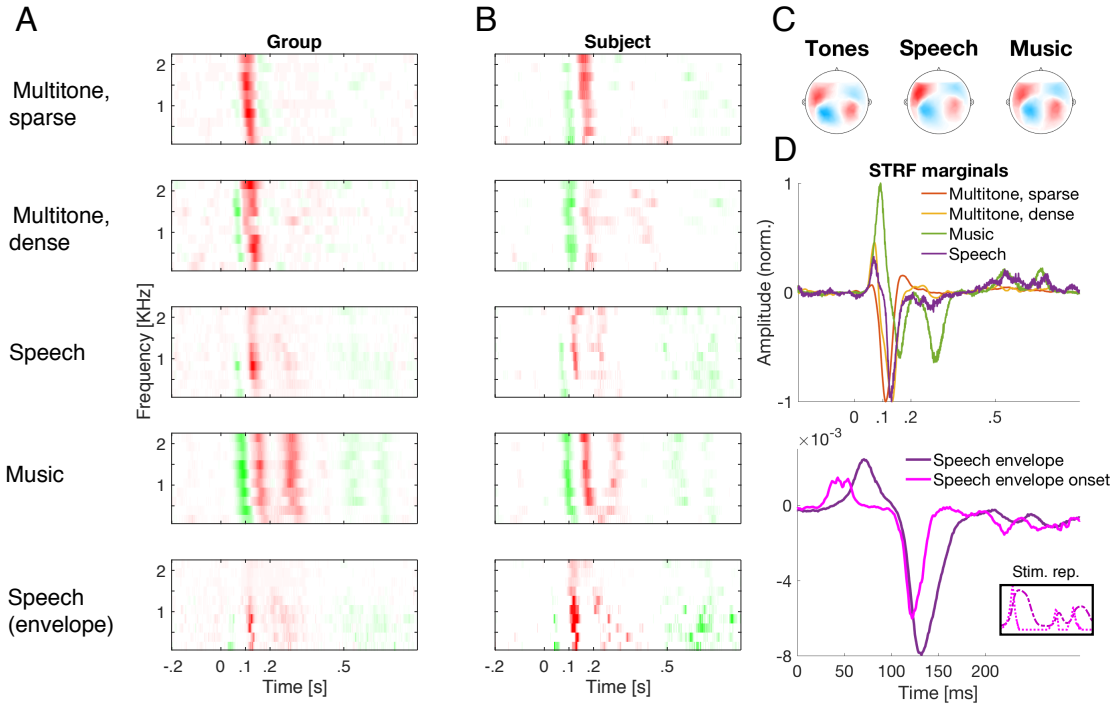

Fig 4. Interpretational power from stimulus representations across STRFs from different stimulus classes. a) Group normalized STRFs from the multitone pattern experiment (*N*=15), and from studies on natural speech (*N*=12) and on music (*N*=15), reveal considerable structural similarity when stimulus onset is extracted as a driving feature of the neuromagnetic response. b) Neuromagnetic STRFs from the same participant across the tones, speech, and music studies, which show substantial consistencies across stimuli when represented by their temporal envelope onsets per frequency band. c) The topographic distribution from same subject as in (b) revealed strong bilateral consistency across classes but with increased left hemisphere-bias during speech processing. d) Top: Timing of major neuromagnetic activity peaks, as shown by TRFs derived from spectral integration of the STRFs in (a), results vary depending on stimulus class and/or context: earliest positive and negative deflections change with increasing acoustic density but also with additional spectrotemporal complexity as found in natural speech and music. Bottom: Group TRFs comparing both speech envelope and envelope onset related activity. Timing differences are explainable by differential acoustic representation in early (< 0.1 s) but not late activity peaks, suggesting the formation of higher order neural representation of elements in speech acoustics by ~100 ms. Only the first deflection timing difference is explained by slope-to-maximum time differences between stimulus representations (inset, same color coding). Curves smoothed by a 5 ms moving average.

To better illustrate class-dependent temporal differences across the studies, TRFs were obtained by collapsing STRFs across spectral bins, as shown in Fig 4d. These plots

emphasize spectrally consistent changes in temporal processing due to stimulus class, along with relative amplitude differences. As before, early activity appeared least prominent for the spectrotemporally sparsest stimuli; in the case of the single participant tested across all three stimulus classes, a high-temporal resolution analysis of the multitone TRF_{100} shows its dynamics are very close to those of the speech envelope counterpart (S3a Fig), with characteristic time constants of ~3 ms (S3b Fig). The response dynamics for music, however, do not follow similarly, which suggests that features other than overall acoustic onsets may contribute to synchronized auditory responses in these cortical populations.

Cortical transformation of natural speech envelope representation

In reverse correlation analyses, exploration of alternative representations of the stimulus may provide complementary insight into the functional operations by the auditory system. Fig 4a and Fig 4b show that for natural speech, STRFs based on the acoustic envelope (row 5) led to functionally informative STRFs, consistent with prior approaches[36,37]. STRFs based on the envelope *onset* representation (row 3) are similar, which is expected since the envelope onset is correlated with the original envelope. In terms of timing, the corresponding group TRFs (Fig 4d) show a difference of 43 ms between TRF_{50} peak components. This was found to be the same as the

characteristic delay between their underlying representations, obtained by cross-correlation of the stimulus representations. Such a close correspondence is evidence that at the level of the neural source of the TRF_{50}, an increasing acoustic envelope operates as a fundamental auditory feature of the stimulus. In contrast, the corresponding comparison of $STRF_{100}$ peaks across the two representations (envelope and envelope onset) gives a much reduced difference of 20 ms (Fig 4d, S5 Fig), not consistent with the acoustic differences between the corresponding representation peaks. Compression in components' relative delays were observed across spectral bins (S2b Fig), as well as in individual temporal response functions (S3a Fig, S4 Fig).

Discussion

The present investigation describes STRFs as a series of response function mappings from artificial and natural sounds to auditory neural responses. It has been demonstrated that these STRFs possess similar predictive power as their single-unit cortical counterparts, and, importantly, show strong similarities across stimulus classes when an acoustic envelope onset representation is chosen. Specific choices of spectrotemporal stimulus representations[14] result in STRF models that are not only predictive but whose temporal structure is highly consistent with that from standard evoked potential components.

Comparison to spike-based spectrotemporal receptive fields

The spectrotemporal *receptive field* can be considered a spike-triggered averaged spectrogram, from auditory periphery[6,38] or central nervous system recordings[2,39–41]. Since reverse correlation is a more general principle than spike-triggered averaging[42] it has been used here to characterize and predict the neural responses of auditory systems where both input and output are continuous time-series[43] via the underlying *response function* of the system. Whether measured by spikes or continuous neural responses, neural systems are non-linear, so predictive linear models of central neural coding are necessarily incomplete descriptions of the underlying coding relationship and are bounded by the predictive power and interpretability they maintain within the limits of the linear regime[11].

The multitone stimulus employed here is comparable to a dynamic random chord stimulus[11,44,45] though it has more temporal degrees of freedom, allowing cross-frequency overlap in a continuing pattern that prevents constant tone presentation rates. It is more similar to dynamic random chords than other artificial stimuli used to estimate STRFs, such as ripple noise and moving ripples[46,47], which focus on stimulus modulations instead. The predictable fraction of variance in the evoked MEG source timeseries was found to be 19–27%, in close correspondence with 18%[11] and 31%[48] predictive power from primary auditory cortex (A1) single/multiunit responses. Comparisons regarding predictive power (and other STRF properties) should also take into account fundamental differences in the underlying signal (spiking versus dendritic-

origin activity) and its scale (neuron or highly local population versus meso-scale cortical patches[34,49]), the animal model, and state (e.g., performing a task vs. resting vs. anesthetized)[7].

Qualitatively, the STRFs presented here exhibit a general broadband structure with frequency-dependent latencies and amplitude changes depending on stimulus density. Remarkably, similar properties appear in STRFs obtained from LFP in mammalian A1[18,19,50], featuring broadband inhibitory-excitatory component sequences and, often, frequency-dependent latencies. Component latencies in mammalian A1 are ~50% shorter than here, which may be explained by reduced equivalent cortico-cortical transmission length delays[51] for the species involved in those studies. With respect to human studies, the component latencies reported here are consistent with multiunit activity[52] and high-gamma activity in electrocorticography (ECoG)[45] in the functional equivalent of A1, Heschl's gyrus. The STRFs obtained in such datasets principally reflect neural spiking, resulting in mappings with narrow-band features, consistent with their interpretation as units locally sampled along the tonotopic gradient of A1. Indeed, frequency selectivity becomes reduced for local field potential recordings[18,50] (i.e. ECoG frequencies below high-gamma) as they sample redundant activity across distant recording sites with intra-cortical interactions[53] – which may effectively smooth the spectral selectivity distribution[18]. Unlike local recordings, which due to high-frequency selectivity can require that receptive fields be realigned by best frequency[28] to extract statistical features, MEG STRFs offer distributed access to more global cortical network domains. Plausibly, analog results may be expected from future human auditory LFP

STRF studies from invasive procedures, given that these have typically focused on multiunit and high-gamma activity[45].

In addition, these MEG-based tone-generated STRFs show stimulus-dependent differences as seen elsewhere in the receptive field literature (see review by Eggermont[7]), namely, amplitude decreases with density. This is consistent with awake primate results, where three-fold increases in tone presentation rate (9.7 to 31 tones/s) may be accompanied by a magnitude decrease of about a third in the STRF maxima[12]; here, a similar multiplicative change in tone density (2 to 6 tones/s) produced a peak decrease of about half. Suppression of excitatory contributions[15], or emergent inhibitory activity throughout A1 single units[12] have been proposed as mechanisms for response field modulations observable from LFP recordings[50]. In cortical neurons, increased firing rates may accentuate depression rate imbalances between excitatory synapses and those with increasingly inhibitory activity[12,19,28], which is a known factor involved in receptive field modulations in somatosensory[54] and visual areas[55]. For auditory recordings, more inhibition may effectively increment responses' spectral specificity or bandwidth at higher tone densities[12,50] – the analog of which was not observed in the MEG STRFs (see Westö & May[56] for a cautionary note on interpreting inhibitory contributions to STRFs following dense stimuli). Among factors reducing STRF predictive power is the increase of inhibitory fields in estimates from single unit recordings[12]; in MEG this effect appeared to be mirrored by response function components of opposite sign to the $STRF_{100}$. Further research is thus necessary

concerning the coarse-grained level of analysis that is accessible via MEG/EEG respectively, in comparison to that afforded by single/multiunit signals.

Association with auditory evoked potentials

Unlike traditional averaging methods, reverse correlation involves continuous delivery of a dynamic stimulus in order to generate a predictive model (of novel instances of the same sound class). It has been shown here that STRFs and TRFs can be directly compared to standard auditory evoked responses, namely the magnetic M50, M100, late auditory evoked responses, and the P1-N1-P2 complex in EEG[26]:

(i) The earliest positive-polarity component, the $STRF_{50}$, seen at higher multitone densities, is a temporal analog to the M50 response originating from Heschl's gyrus (including core/primary areas)[57,58]; its amplitude may also be modulated by inter-stimulus intervals[58] at low presentation rates (<2 tones/s). Known modulators of M50 amplitudes include harmonic versus noise-like bursts[59,60], prepulse inhibition[61], and automatic processing of redundant information as a form of sensory gating in paired-click stimulus designs[62,63]. In terms of predictive power, this component did not generalize well over novel instances of the multitone random pattern; this is consistent with an adaptive role contributing to considerable changes in the response profile dependent on the local context on the order of a few seconds or less.

(ii) The subsequent major component, the negative-polarity $STRF_{100}$, exhibited magnitude decrease and delay increase with density, with a sharp transition after the

sparsest density level. Suppression of the M100 response from supratemporal cortex has been observed in the transition from low to higher tone presentation rates[58] highlighting the interpretation of increased inhibitory effects that include generalized refractoriness among neurons at denser conditions[64]. This component is also subject to attentional modulation[65–67] which may reflect that individual tones in a densely populated scene fail to capture attention individually. Because of this component's involvement in tracking perceptual objects of an auditory scene[37], and of the increasing quality of flow and continuity in these artificial stimuli, sharp transitions in this component may suggest indices of 'crowding' relevant to the figure-ground separation problem[68,69]. Accounts of spectrally-dependent latency in the evoked M100 components[20–25] were consistent at this stage and fall within the sensitivity domain for human voice pitch production and discrimination[24]. These latencies are also consistent with those of pitch-specific onset responses, whether elicited by complex tones or by centrally-generated Huggins pitch percepts[70,71].

(iii) The second major positive-polarity peak appears only in response to the sparsest stimulus. Auditory event-related potentials at ~200 ms latency have been described in EEG as expectancy indices, exhibiting greater amplitudes for tones whose presentation in time is uncertain[72]. At denser conditions, shorter inter-stimulus-intervals may reduce the tone-evoked analogous EEG P2 component, regardless of presentation within a repetition sequence or as an oddball, suggesting involvement of modulation mechanisms other than habituation[64].

On alternative representations of stimulus state

In addition to their predictive power, STRF profiles are functionally informative in a way similar to trial-averaged evoked responses to isolated stimuli[8,14,73]; this was the case for STRFs compared across stimulus classes when the stimulus representations were filterbank-derived onsets. Other abstract representations of this stimulus pattern, including both temporal edges, their directionality, the duration of sustained acoustic energy, or, related to the latter, the spectrogram, did not appear to be similarly functionally informative – even though they contained and extended information from onset representation. Nevertheless, predictive power was similar across representations, suggesting that this metric alone is insufficient to expose which aspects of the stimulus map to the system's response. More complex tone patterns might allow predictive power to become more informative regarding the statistical characterization of a stimulus (c.f.[74]). The lack of evidence for explicit neural encoding of offsets is in accord with neurophysiological evidence suggesting offset-encoding cells to be outnumbered by onset cells, and/or to have minor neural response profiles relative to onset encoders[30,31], which in the aggregate would result in differential contributions to the neuromagnetic response.

On extension to natural stimuli

Processing of environmental sounds, including conspecific calls, is a critical auditory task. Encoding models incorporating natural sounds with complex spectrotemporal

structure provide powerful computational insights into the auditory system that may be inaccessible with synthetic stimuli only[7–9]. STRFs derived from invasive recordings from A1 perform similarly in terms of predictive power, using random tone chord stimuli, animal calls, environmental sounds, sound effects, and music[11,27], at the population level. For some subset of these neurons, successful linear encoding of the spectrogram may also occur in the same unit for *both* artificial *and* natural vocalization encodings[10,28]. The search for predictive models that generalize over novel stimuli not in the training set has proven difficult however[10,44]. The temporal statistics intrinsic to natural sounds may be critical[10], and some evidence from A1 STRFs demonstrates higher predictive power using conspecific vocalizations that are not dilated or compressed[75]; similarly, comparisons are also favorable for artificial and communication sounds controlled for the span of their temporal and spectral modulations jointly, but allowing differences in their amplitude fluctuations over time[10]. Observed stimulus-class dependencies in STRF spectrotemporal properties appear as small time-shifts of STRF features, plus the emergence of additional late activity for speech and music. Analysis of such differences suffer from confounds arising from statistical non-uniformities among the sampled classes[8–10,29] and fully addressing this issue is beyond the scope of this investigation. The question of whether detailed class-dependent temporal coding frameworks may be achieved by means of linear methods remains open.

On speech-derived STRFs

Advances in understanding cognitive processes relevant to speech processing have

followed from reverse correlation studies that used the speech acoustic envelope (as represented by low-frequency, 1-15 Hz fluctuations in ECoG[76,77] and MEG/EEG[36,37,78,79] recordings). We find that the speech envelope $STRF_{100}$ component exhibits similar spectral-dependent latency as the M100 evoked response, thus suggesting a level of speech analysis that still contains independent spectral information. Although in contrast with findings of near-constant M100 latencies for certain synthetic vowels presented in isolation[24], reverse correlation methods over long natural speech presentations are better suited to probe domain-general processing in realistic conditions due to their extended sampling,.

Additionally, the methods may constrain the time course of the change in neural representations of human speech from spectrogram to higher level. The low-frequency speech envelope and its onsets are both operationally related to functionally informative STRFs. The systematic delay between timeseries (peaks in the latter systematically precede those in the former) directly accounted for the relative difference between the resulting pair of $STRF_{50}$ components after each representation. The acoustic mismatch could not explain, however, the reduced relative difference between subsequent $STRF_{100}$ pairs. The interpretation of a compression is consistent with current models of step-wise speech processing, where the formation of speech analysis units or objects is preceded by an earlier spectrogram-like representation of acoustics completed by ~80 ms post speech impulse onset. After this time, response functions did not account for the expected mismatch, suggesting a neurally-based progression into a modified stage of the neural representation of speech and adding to a body of MEG evidence for a cortical hierarchy of speech object representations (see review by Zhang and colleagues [80]).

Overall, these results demonstrate an important advantage of STRFs over standard epoch-averaging methods commonly when used in MEG applications, e.g., characterizing the phenomenology of disorders in clinical populations[81]: their ability to generalize to critical sounds beyond pure tones, most importantly natural speech. By providing both neural predictions and functional information, it allows noninvasive approaches to understanding developmental[82], learning and associative effects induced by tasks[83–85], or behavioral contexts[86,87] – thus potentially furthering insight into the role of dynamical representations of sound in auditory cognition.

Methods

Participants

15 subjects (6 women, 23.2 ± 2.9 years of age [mean ± SD]), 1 left-handed[88], participated in the multitone study. 12 subjects (6 women, 24.1 ± 3.0 years of age), all right-handed native English speakers, participated in the speech study. 15 subjects (5 women, 21.0 ± 1.7 years of age), all right-handed, participated in the music study. Each subject received monetary compensation proportional to the study duration (approximately 1.5 hours). Subjects had no history of neurological disorder or metal implants. The experimental protocol was approved by the UMCP Institutional Review Board and before each study session, informed written consent was obtained from the participant.

Stimuli

Multitone study. Sound stimuli were constructed with the MATLAB® software package (MathWorks, Natick, United States) at a sampling rate of 44.1 KHz, and consisted of 50 s auditory scenes composed of pseudo-randomly presented 180 ms tones, each with frequency f_i taken from a pool of 10 fixed values (range: 180-2144 Hz) in 2 equivalent rectangular bandwidth (ERB) steps[89] specified by $f_i = f_{i-1} + \left(24.7(1 + 4.37 f_{i-1}/1000)\right)$. For each frequency, tone onset times were uniformly distributed with a minimum inter-tone gap of 40 ms. Five tone presentation rates (2, 4, 6, 8, or 10 per second over all channels) were used separately. Tone onset times T_j were independent across frequency bands and selected in 20 ms bins. Individual tones were modulated with 10 ms raised cosine on- and off-ramps. Tone level was calibrated according to frequency based on the 60-phon normal equal-loudness-level contour (ISO 226:2003) in order to adjust for perceived relative loudness differences; relative gains to a 1 KHz reference were determined in 2 dB SPL steps. *Speech and music studies*. For the speech study, a 60 s female voice audiobook excerpt [90] narrated from *The Light Princess* (Macdonald, 1864) was used as part of a related study on reverberant speech processing [16]. For the music study, 55 s samples across 6 different instrumental musical styles reflecting a variety of genres and traditions, were presented: orchestra, *Symphony in F Major, No. 32, Movement I* (Sammartini, c. 1740); swing, *Cascades* (Combelle, c. 1940); blues, *Blues for B&W* (Rogers & Hilden, 2003); sarangi, *Raga Mishra Bhairavi: Alap* (Narayan, 2002); pipa, *Dance of the Yi People* (Huiran, c. 1960); and a euphonium transcription of

Dancing Night Wind (Benning, 1997).

In all studies, audio signals were normalized and presented through the Presentation® software package (NeuroBehavioral Systems, Berkeley, United States), using audio equipment equalized to a transfer function approximately flat from 40 to 3000 Hz. Sound stimuli were transmitted to subjects via ear insertion tubes E-A-RTONE® 3A of 50 Ω impedance and E-A-RLINK® disposable foam intra-auricular ends (Etymotic Research, Elk Grove Village, United States) that were inserted in the ear canals.

Experimental design

For the *multitone* study, trials consisted of a main tone cloud pattern scene presented in series with per block, generated anew per each subject. This resulted in trials that contained between 0 and 3 multitone density transitions within the trial, and ranged from 70 to 120 s duration. Each of the five main scenes were repeated 4 times, and only these data epochs were analyzed. After a brief training session, subjects were instructed to attend to the ongoing stimulus with their eyes closed and to report rate transitions via a button press. Optional rests were available every 5 trials, totaling 1.5 hours recording time. Subjects received feedback on the correct number of transitions at the end of each trial. For the *speech* study, trials consisted of various story passages presented in random order at different reverberant noise levels. At the end of a trial, subjects were asked comprehensive questions about the passage, and rated its intelligibility. For the present study purposes, analysis was based exclusively on reverberation-free, no-noise ('clean') trials, repeated 3 times across the experiment. For the *music* study, trials consisted of

each of the 6 samples presented individually in random order. At the end of each trial, a 5 s clip taken from the same or a different piece was presented and subjects identified if it was an excerpt of the preceding trial. Each sample repeated 3 times across the experiment.

Neural data recording

Magnetoencephalography (MEG) data were collected with a 160-channel system (Kanazawa Technology Institute, Kanazawa, Japan)[91] inside a magnetically-shielded room (Yokogawa Electric Corporation, Musashino, Japan) at a sampling rate of 1 KHz. Superconducting quantum interference device (SQUID) sensors (15.5 mm diameter each) were uniformly distributed (~25 mm) inside a Dewar vase containing liquid-He refrigerant, with a concave outer surface fit to the average human head. Sensors are first-order axial gradiometers with 50 mm separation and sensitivity greater than 5 $fT \cdot Hz^{-1/2}$ in the white noise spectral region (> 1 KHz), except for three additional reference magnetometers separated from the neural sensors and arranged orthogonally to each other. A 1 Hz high-pass analog filter, a 200 Hz low-pass analog filter, and a 60 Hz analog notch filter were applied online respectively. Sensor channels with saturating or zero responses over more than 12.5 s recording time were excluded from analysis. Participants laid supine inside the magnetically shielded room and were asked to minimize body movement, particularly from the head.

Neural data processing

Environmental noise. To eliminate environmental magnetic noise contributions, time-shifted principal component analysis[92] (TS-PCA) was applied, a process that discards optimally-filtered environmental signals recorded on the reference sensors. Reference sensors were 3 physical magnetometers (see *Neural Data Recording*) plus 2 virtual channels obtained by independent component analysis[93] of the remaining data sensors and selecting the two components with the most unstructured broadband (0-500 Hz) power. *Sensor-specific noise*. Electronic sensor noise was removed via sensor noise suppression (SNS)[94] by substituting each channel signal with its projection onto the orthogonal basis space generated by all other sensors in the system. This method exploits redundant activity across elements of the dense array (where the number of channels exceeds the number of brain sources of interest) by attenuating components specific to any single channel. *Spatial filtering*. Data-driven spatial filters were derived per participant using responses evoked by repeated trials in each of the respective studies. Response epochs of 45-55 s duration were extracted, band-pass filtered (1-15) Hz with a 2nd order Butterworth filter, and delay corrected (~13 ms). A linear transformation based on this manipulation was obtained per participant[95] to generate spatial filters that correspond to magnetic fields generated by the left and right auditory cortex (S3 Fig). This spatial filter was applied to the raw data and the resulting neural signal, representing the most reproducible component of the evoked data, was selected as a single virtual sensor in analyses henceforth.

Neural data analysis

Spectrotemporal response function of stimulus representation. For multitone patterns, pure onset representations only carry information at a time beginning with the onset of a tone. We formulate this representation as

$$O(f,t) = \delta_{t-T_{ij}}\delta_{f-f_i} \tag{1}$$

where every onset has equal weight independent of its tone's frequency band f_i (i = 1,…,10), with onset timesT_{ij} T_{ij} of the j-th tone with frequency f_i; δ_n is the discrete unit impulse centered at sample n. The input-output relation between this representation of auditory input and the evoked cortical response $\bar{r}(t)$ is then modeled by a spectrotemporal response function (STRF). For discrete data this linear model is formulated as:

$$r^*_{pred}(t) = \sum_f \sum_\tau STRF(f,t)O(f,t-\tau) + \varepsilon(t) \tag{2}$$

where $\varepsilon(t)$ is the residual contribution to the evoked response not explained by the linear system. Summing only over the frequency term allows evaluating the temporal profile of the response function model (TRF). Exploration of alternative stimulus representations requires substitution of the $O(f,t)$ term in (2) by the analogous time-frequency representation of the stimulus (e.g. by a spectrogram $S(f,t)$). For all stimuli, *stimulus envelope filterbank* representations were obtained by passing the original waveform through a filterbank of ten order 1000 FIR filters with passbands at mid-values between f_i neighbors (see *Stimuli*, above) starting at 143 Hz. Filter delays were compensated and the envelope in each band was extracted as above. Sampling rates were reduced to 1 KHz

and signals smoothed by a delay-corrected 4th order binomial FIR filter. Half-wave rectification (i.e. setting negative values to zero) of the derivative of the stimulus envelope filterbank output gave *envelope onset* representations of the stimulus signal based on the filterbank. Prior to reverse correlation, both envelope and envelope onset representations were transformed to dB-scale.

Linear STRF model estimation. STRF estimation was performed via *boosting*, a technique where the error estimate $\varepsilon(t)$ (in Eq. 2) is minimized iteratively via sequential modifications to the STRF[29]. The name originates from the ability to improve ('boost') an estimate learning algorithm by establishing aggregate decision rules from across a sequence of many estimation steps, each needing only slightly-better-than-chance accuracy[96–98]. This technique can then be implemented as a forward stage-wise fitting that follows a greedy heuristic, by adding the contribution with the largest available mean-squared-error reduction at each given step[29,99] and in turn maximizing the predictive power of the model[11]. Operationally, STRF estimates by boosting were initialized as a null matrix of dimensions TxF, where T equals the number of experimental time bins and F is the total of frequency bins (=10; for TRF estimates, F=1); optimization followed through exploring fixed increments and decrements per spectrotemporal bin individually. Among the resulting 2xFxT possible choices, the outcome with minimum mean-squared-error was selected as the next step in the running STRF estimate. The procedure was iterated, accumulating optimizations, until modifications instead produced a sustained increase in mean-squared error[99], since the method is not guaranteed to find a global optimum. This termination method effectively imposes a sparse structure on the STRF, which allows for extraction of high-temporal

resolution features in the STRF even if only low-frequency content was present in the input waveforms (other STRF estimation methods such as normalized reverse correlation[8] and generalized linear models could also used [29,100,101]). Other detailed descriptions of the boosting algorithm implementation for timeseries data, including MEG/EEG are available[29,36].

STRF predictive power bounds. The measured evoked cortical response $r^*(t)$ may include stimulus-independent noise, the presence of which is a consequence of the finite dataset size and leads to STRF model parameters that overfit to the training data. Performance measures that account for stimulus-independent noise are necessarily overestimates and therefore can be considered to act as empirical upper bounds of model performance[11]. In contrast, the risk of overfitting can be minimized using *cross-validation*, where a fraction of the $r^*(t)$ timeseries $\bar{r}(t)$ (e.g. 90%) is reserved for model training, and testing is done on the remaining fraction incorporating only the model's ability to generalize over novel stimulus instances. This would be expected to underperform with respect to an optimal model for the dataset in question and so indicates a lower bound for its performance [11]. In practice, it is this conservative, cross-validated lower-bound that is used for STRF estimates.

Nonlinear extension. Linear encoding models may fail to characterize firing rate predictions based on effects such as threshold activity, past-history dependencies, dynamic range compression, synaptic transfer, and the non-negative distribution of the neuron response for example. At the single neuron level, predictions can be improved via introduction of static nonlinearities derived by empirical fit[42], or via intermediate nonlinearities in more complex model hierarchies[101]. For coarse-grained continuous

neural responses such as local field potentials and the MEG signal here, it appears that such model hierarchies may no longer apply well. When a static nonlinearity was incorporated using a linear-nonlinear (LN) model[102,103], only a 2% improvement to predictive power resulted (quadratic fit, R^2=0.972, S5 Fig) and so was not pursued.

Estimation of STRF predictive power and noise limit extrapolation. To assess STRF model validity, predictive power was estimated as the fraction of a response signal variance that is stimulus-explained, and corrected for the reduction of noise-related variance achieved by averaging[11]. Namely, for MEG response timeseries $r_1(t)$, … ,$r_N(t)$$\mathrm{r}_1(\mathrm{t}), \ldots, \mathrm{r}_\mathrm{N}(\mathrm{t})$ where N is the number of repetition trials, total variance is expressed as the average of each trial's individual variance

$$\mathrm{Var}(r) = \frac{1}{N}(\mathrm{Var}(r_1(t)) + \cdots + \mathrm{Var}(r_N(\mathrm{t}))) \tag{3}$$

while evoked variance can be expressed as that of the average response $\mathrm{Var}(\bar{r})$. When N is large, the extent to which total variance is larger than evoked variance indexes reliability for the response source. Contributions to total variance $\mathrm{Var}(r)$ are then partitioned into those stemming from the evoked *signal*, and the remainder is treated as *noise*:

$$\mathrm{Var(signal)} = \frac{1}{N-1}\left(N \cdot \mathrm{Var}(\bar{r}) - \mathrm{Var}(r)\right) \tag{4}$$

$$\mathrm{Var(noise)} = \frac{N}{N-1}\left(\mathrm{Var}(r) - \mathrm{Var}(\bar{r})\right) \tag{5}$$

such that estimates are corrected for cases where N is small. Often, STRF model estimates are optimized to produce accurate predictions of the evoked response only; in such cases, use of single-trial variance provides an additional statistic regarding the event-related contribution to available recordings. Once a STRF model has been obtained

for a particular condition and subject, its ability to predict the evoked response is assessed as the extent of evoked response variance that is not residual error, that is $\mathrm{Var}(\bar{r}) - \mathrm{Var}(\bar{r} - \bar{r}_{pred})$. This expression is the model's predictive power, which after division by the estimated signal power (eq. 4)$\mathrm{Var}(\mathrm{signal})$, represents the fraction of stimulus-evoked variance described by the linear STRF model contingent on a given experimental condition and subject. Analogously, noise power in the same response may be normalized by the estimated signal power, providing the inverse proportion to which the procedure of averaging reduces response variability. When *N* is very large, a normalized noise power of e.g. 10 indicates that averaging reduces variance in the evoked signal to almost a tenth of the original total variance. In the hypothetical case where the procedure of averaging yields no reduction in variability (such as with identical trial response instances), the absence of variability reduction implies an absolute zero noise level. Empirically, each dataset's (condition and subject) predictive power can be indexed by the intrinsic noise power (e.g. Fig 1C). Assuming the responses have been measured from a similar population, regression analysis may produce an estimate of the STRF model class predictive power, via its extrapolation to the theoretical noise-free limit[11].

Acknowledgements

We thank Elizabeth Nguyen for excellent technical assistance. This study was funded by the US National Institutes of Health (R01-DC-00843, R01-DC-014085) and the US National Science Foundation (OAC-1063035). We thank support to MVD by the NSF (DGE-142487) and to FCC by the Mexican Consejo Nacional de Ciencia y Tecnología through its graduate scholarship program. The funders had no role in study

design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Woolley SMN, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci. 2005;8(10):1371–9.

2. Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J Neurophysiol. 2001;85(3):1220–34.

3. Kumar S, Penny W. Estimating neural response functions from fMRI. Front Neuroinformatics. 2014;8:48.

4. Aertsen AMHJ, Johannesma PIM. The Spectro-Temporal Receptive Field. Biol Cybern. 1981;42(2):133–43.

5. Aertsen AMHJ, Johannesma PIM. Spectro-temporal receptive fields of auditory neurons in the grassfrog. Biol Cybern. 1980;38(4):223–34.

6. Eggermont JJ, Johannesma PIM, Aertsen AMHJ. Reverse-correlation methods in auditory research. Q Rev Biophys. 1983;16(03):341–414.

7. Eggermont JJ. Context dependence of spectro-temporal receptive fields with implications for neural coding. Hear Res. 2011;271(1–2):123–32.

8. Theunissen FE, Sen K, Doupe AJ. Spectral-Temporal Receptive Fields of Nonlinear Auditory Neurons Obtained Using Natural Sounds. J Neurosci. 2000;20(6):2315–31.

9. Theunissen FE, Elie JE. Neural processing of natural sounds. Nat Rev Neurosci. 2014;15(6):355–66.

10. Laudanski J, Edeline J-M, Huetz C. Differences between spectro-temporal receptive fields derived from artificial and natural stimuli in the auditory cortex. PloS One. 2012;7(11):e50539.

11. Sahani M, Linden JF. How Linear are Auditory Cortical Responses? In: Becker S, Thrun S, Obermayer K, editors. Advances in Neural Information Processing Systems 15. MIT Press; 2003. p. 125–132.

12. Blake DT, Merzenich MM. Changes of AI Receptive Fields With Sound Density. J Neurophysiol. 2002;88(6):3409–20.

13. Meyer AF, Williamson RS, Linden JF, Sahani M. Models of Neuronal Stimulus-Response Functions: Elaboration, Estimation, and Evaluation. Front Syst Neurosci. 2017;10.

14. Gill P, Zhang J, Woolley SMN, Fremouw T, Theunissen FE. Sound representation methods for spectro-temporal receptive field estimation. J Comput Neurosci. 2006;21(1):5–20.

15. Valentine PA, Eggermont JJ. Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. Hear Res. 2004;196(1–2):119–33.

16. Villafañe-Delgado M. The Cortical Representations of Speech in Reverberant Conditions MS Thesis. University of Maryland, College Park; 2013. Available from: http://drum.lib.umd.edu/handle/1903/14654

17. Camenga E, Dombrowski K, Walsh B, Cervantes Constantino F, Puvvada K, Villafañe-Delgado M, et al. Cortical Representations of Music in Human Listeners. Poster presented at: Midwinter Meeting of the Association for Research in Otolaryngology; 2013; Baltimore, MD

18. Gaucher Q, Edeline J-M, Gourévitch B. How different are the local field potentials and spiking activities? Insights from multi-electrodes arrays. J Physiol-Paris. 2012;106(3–4):93–103.

19. Gourévitch B, Noreña A, Shaw G, Eggermont JJ. Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent. Cereb Cortex; 1991;19(6):1448–61.

20. Roberts TP, Poeppel D. Latency of auditory evoked M100 as a function of tone frequency. Neuroreport. 1996;7(6):1138–40.

21. Greenberg S, Poeppel D, Roberts T. A Space-Time Theory of Pitch and Timbre Based on Cortical Expansion of the Cochlear Traveling Wave Delay. 1997. In A. Palmer, Q. Summerfield, A. Rees, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 293-300). London: Whurr Publishers.

22. Lütkenhöner B, Steinsträter O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol Neurootol. 1998;3(2–3):191–213.

23. Roberts TP, Ferrari P, Stufflebeam SM, Poeppel D. Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. J Clin Neurophysiol Soc. 2000;17(2):114–29.

24. Mäkelä AM, Alku P, Mäkinen V, Valtonen J, May P, Tiitinen H. Human Cortical Dynamics Determined by Speech Fundamental Frequency. NeuroImage. 2002;17(3):1300–5.

25. Salajegheh A, Link A, Elster C, Burghoff M, Sander T, Trahms L, et al. Systematic latency variation of the auditory evoked M100: from average to single-trial data. NeuroImage. 2004;23(1):288–95.

26. Picton TW. Human auditory evoked potentials. San Diego: Plural Pub; 2010. 634 p.

27. Machens CK, Wehr MS, Zador AM. Linearity of cortical receptive fields measured with natural sounds. J Neurosci. 2004;24(5):1089–100.

28. David SV, Mesgarani N, Fritz JB, Shamma SA. Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. J Neurosci Off J Soc Neurosci. 2009;29(11):3374–86.

29. David SV, Mesgarani N, Shamma SA. Estimating sparse spectro-temporal receptive fields with natural stimuli. Netw Comput Neural Syst. 2007;18(3):191–212.

30. Recanzone GH. Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. Hear Res. 2000;150(1–2):104–18.

31. Phillips DP, Hall SE, Boehnke SE. Central auditory onset responses, and temporal asymmetries in auditory perception. Hear Res. 2002;167(1–2):192–205.

32. Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by preferred stimuli. Nature. 2005;435(7040):341–6.

33. Qin L, Chimoto S, Sakai M, Wang J, Sato Y. Comparison between offset and onset responses of primary auditory cortex ON-OFF neurons in awake cats. J Neurophysiol. 2007;97(5):3421–31.

34. Lopes da Silva FH. Electrophysiological Basis of MEG Signals. In: Hansen P, Kringelbach M, Salmelin R, editors. MEG: An Introduction to Methods. Oxford University Press; 2010. p. 1–23.

35. Power AJ, Reilly RB, Lalor EC. Comparing linear and quadratic models of the human auditory system using EEG. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2011;2011:4171–4.

36. Ding N, Simon JZ. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol. 2012;107(1):78–89.

37. Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci. 2012;109(29):11854–9.

38. de Boer E, Kuyper P. Triggered Correlation. IEEE Trans Biomed Eng. 1968;BME-15(3):169–79.

39. Miller LM, Escabí MA, Schreiner CE. Feature selectivity and interneuronal cooperation in the thalamocortical system. J Neurosci. 2001;21(20):8136–44.

40. deCharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons. Science. 1998;280(5368):1439–43.

41. Escabi MA, Schreiner CE. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. J Neurosci. 2002;22(10):4114–31.

42. Dayan P, Abbott LF. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. The MIT Press; 2005.

43. Eggermont JJ. Wiener and Volterra analyses applied to the auditory system. Hear Res. 1993;66(2):177–201.

44. Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I. Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. Nature. 2008;451(7175):197–201.

45. Jenison RL, Reale RA, Armstrong AL, Oya H, Kawasaki H, Howard MA. Sparse Spectro-Temporal Receptive Fields Based on Multi-Unit and High-Gamma Responses in Human Auditory Cortex. PloS One. 2015;10(9):e0137915.

46. Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust Spectrotemporal Reverse Correlation for the Auditory System: Optimizing Stimulus Design. J Comput Neurosci. 2000;9(1):85–111.

47. Escabí MA, Read HL. Representation of spectrotemporal sound information in the ascending auditory pathway. Biol Cybern. 2003;89(5):350–62.

48. Williamson RS, Ahrens MB, Linden JF, Sahani M. Input-Specific Gain Modulation by Local Sensory Context Shapes Cortical and Thalamic Responses to Complex Sounds. Neuron. 2016;91(2):467–81.

49. Silva FL da. EEG: Origin and Measurement. In: Mulert C, Lemieux L, editors. EEG - fMRI. Springer Berlin Heidelberg; 2009. p. 19–38.

50. Noreña AJ, Gourévitch B, Pienkowski M, Shaw G, Eggermont JJ. Increasing spectrotemporal sound density reveals an octave-based organization in cat primary auditory cortex. J Neurosci. 2008;28(36):8885–96.

51. Wang SS-H, Shultz JR, Burish MJ, Harrison KH, Hof PR, Towns LC, et al. Functional trade-offs in white matter axonal scaling. J Neurosci. 2008;28(15):4047–56.

52. Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M. Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. J Neurosci. 2011;31(40):14067–75.

53. Kaur S, Lazar R, Metherate R. Intracortical pathways determine breadth of subthreshold frequency receptive fields in primary auditory cortex. J Neurophysiol. 2004;91(6):2551–67.

54. Galarreta M, Hestrin S. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. Nat Neurosci. 1998;1(7):587–94.

55. Varela JA, Song S, Turrigiano GG, Nelson SB. Differential depression at excitatory and inhibitory synapses in visual cortex. J Neurosci. 1999;19(11):4293–304.

56. Westö J, May PJC. Capturing contextual effects in spectro-temporal receptive fields. Hear Res. 2016;339:195–210.

57. Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol Potentials Sect. 1994;92(3):204–14.

58. Onitsuka T, Ninomiya H, Sato E, Yamamoto T, Tashiro N. The effect of interstimulus intervals and between-block rests on the auditory evoked potential and magnetic field: is the auditory P50 in humans an overlapping potential? Clin Neurophysiol Off J Int Fed Clin Neurophysiol. 2000;111(2):237–45.

59. Hertrich I, Mathiak K, Lutzenberger W, Ackermann H. Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields. Neuroreport. 2000;11(18):4017–20.

60. Chait M, Simon JZ, Poeppel D. Auditory M50 and M100 responses to broadband noise: functional implications. Neuroreport. 2004;15(16):2455–8.

61. Braff DL, Geyer MA, Swerdlow NR. Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. Psychopharmacology (Berl). 2001;156(2–3):234–58.

62. Grunwald T, Boutros NN, Pezer N, von Oertzen J, Fernández G, Schaller C, et al. Neuronal substrates of sensory gating within the human brain. Biol Psychiatry. 2003;53(6):511–9.

63. Korzyukov O, Pflieger ME, Wagner M, Bowyer SM, Rosburg T, Sundaresan K, et al. Generators of the intracranial P50 response in auditory sensory gating. NeuroImage. 2007;35(2):814–26.

64. Pereira DR, Cardoso S, Ferreira-Santos F, Fernandes C, Cunha-Reis C, Paiva TO, et al. Effects of inter-stimulus interval (ISI) duration on the N1 and P2 components of the auditory event-related potential. Int J Psychophysiol. 2014;94(3):311–8.

65. Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical Signs of Selective Attention in the Human Brain. Science. 1973;182(4108):177–80.

66. Schwent VL, Hillyard SA, Galambos R. Selective attention and the auditory vertex potential. I. Effects of stimulus delivery rate. Electroencephalogr Clin Neurophysiol. 1976;40(6):604–14.

67. Fujiwara N, Nagamine T, Imai M, Tanaka T, Shibasaki H. Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. Cogn Brain Res. 1998;7(2):99–109.

68. Durlach NI, Mason CR, Shinn-Cunningham BG, Arbogast TL, Colburn HS, Jr GK. Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J Acoust Soc Am. 2003;114(1):368–79.

69. Lutfi RA, Gilbertson L, Heo I, Chang A-C, Stamas J. The information-divergence hypothesis of informational masking. J Acoust Soc Am. 2013;134(3):2160–70.

70. Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lütkenhöner B. Neuromagnetic Evidence for a Pitch Processing Center in Heschl's Gyrus. Cereb Cortex. 2003;13(7):765–72.

71. Chait M, Poeppel D, Simon JZ. Neural Response Correlates of Detection of Monaurally and Binaurally Created Pitches in Humans. Cereb Cortex. 2006;16(6):835–48.

72. Sowman PF, Kuusik A, Johnson BW. Self-initiation and temporal cueing of monaural tones reduce the auditory N1 and P2. Exp Brain Res. 2012;222(1–2):149–57.

73. Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA. Temporal Symmetry in Primary Auditory Cortex: Implications for Cortical Connectivity. Neural Comput. 2007;19(3):583–638.

74. Sharpee T, Rust NC, Bialek W. Analyzing neural responses to natural signals: maximally informative dimensions. Neural Comput. 2004;16(2):223–50.

75. Carruthers IM, Natan RG, Geffen MN. Encoding of ultrasonic vocalizations in the auditory cortex. J Neurophysiol. 2013;109(7):1912–27.

76. Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, et al. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." Neuron. 2013;77(5):980–91.

77. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012;485(7397):233–6.

78. Di Liberto GM, O'Sullivan JA, Lalor EC. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr Biol. 2015;25(19):2457–65.

79. Presacco A, Simon JZ, Anderson S. Effect of informational content of noise on speech representation in the aging midbrain and cortex. J Neurophysiol. 2016;jn.00373.2016.

80. Simon JZ. The encoding of auditory objects in auditory cortex: Insights from magnetoencephalography. Int J Psychophysiol. 2015;95(2):184–90.

81. Roberts TPL, Khan SY, Rey M, Monroe JF, Cannon K, Blaskey L, et al. MEG Detection of Delayed Auditory Evoked Responses in Autism Spectrum Disorders: Towards an Imaging Biomarker for Autism. Autism Res Res. 2010;3(1):8–18.

82. Zhang LI, Bao S, Merzenich MM. Persistent and specific influences of early acoustic environments on primary auditory cortex. Nat Neurosci. 2001;4(11):1123–30.

83. Delgutte B. Two-tone rate suppression in auditory-nerve fibers: dependence on suppressor frequency and level. Hear Res. 1990;49(1–3):225–46.

84. Noreña AJ, Gourévitch B, Aizawa N, Eggermont JJ. Spectrally enhanced acoustic environment disrupts frequency representation in cat auditory cortex. Nat Neurosci. 2006;9(7):932–9.

85. Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci. 2003;6(11):1216–23.

86. Weinberger NM. Dynamic Regulation of Receptive Fields and Maps in the Adult Sensory Cortex. Annu Rev Neurosci. 1995;18(1):129–58.

87. Ma L, Micheyl C, Yin P, Oxenham AJ, Shamma SA. Behavioral measures of auditory streaming in ferrets (Mustela putorius). J Comp Psychol. 2010;124(3):317–30.

88. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971;9(1):97–113.

89. Glasberg BR, Moore BCJ. Derivation of auditory filter shapes from notched-noise data. Hear Res. 1990;47(1–2):103–38.

90. Database: LibriVox [Internet]. Available from: https://librivox.org/the-light-princess-by-george-macdonald

91. Kado H, Higuchi M, Shimogawara M, Haruta Y, Adachi Y, Kawai J, et al. Magnetoencephalogram systems developed at KIT. IEEE Trans Appl Supercond. 1999;9(2):4057–62.

92. de Cheveigné A, Simon JZ. Denoising based on time-shift PCA. J Neurosci Methods. 2007;165(2):297–305.

93. Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw. 2000;13(4–5):411–30.

94. de Cheveigné A, Simon JZ. Sensor noise suppression. J Neurosci Methods. 2008;168(1):195–202.

95. de Cheveigné A, Simon JZ. Denoising based on spatial filtering. J Neurosci Methods. 2008;171(2):331–9.

96. Schapire RE. The strength of weak learnability. Mach Learn. 5(2):197–227.

97. Freund Y. Boosting a Weak Learning Algorithm by Majority. Inf Comput. 1995;121(2):256–85.

98. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). Ann Stat. 2000;28(2):337–407.

99. Zhang T, Yu B. Boosting with Early Stopping: Convergence and Consistency. Ann Stat. 2005;33(4):1538–79.

100. Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SMN. A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PloS One. 2011;6(1):e16104.

101. Schinkel-Bielefeld N, David SV, Shamma SA, Butts DA. Inferring the role of inhibition in auditory processing of complex natural stimuli. J Neurophysiol. 2012;107(12):3296–307.

102. Chichilnisky EJ. A simple white noise analysis of neuronal light responses. Netw Comput Neural Syst. 2001;12(2):199–213.

103. Sharpee TO, Miller KD, Stryker MP. On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. J Neurophysiol. 2008;99(5):2496–509.