The Journal of Neuroscience

https://jneurosci.msubmit.net

JN-RM-2107-17R2

Dual coding of frequency modulation in the ventral cochlear nucleus

Nihaad Paraouty, Laboratoire des Systèmes Perceptifs CNRS UMR 8248, Ecole normale supérieure, Paris Sciences et Lettres Research University
Arkadiusz Stasiak, University of Cambridge
Christian Lorenzi, Ecole normale superieure, Paris Sciences et Lettres
Leo Varnet, Ecole normale superieure, Paris Sciences et Lettres
Ian. M. Winter, University of Cambridge

Commercial Interest:

DUAL CODING OF FREQUENCY MODULATION IN THE VENTRAL COCHLEAR NUCLEUS

Nihaad Paraouty[1, 2*], Arkadiusz Stasiak[1], Christian Lorenzi[2], Leo Varnet[2] and Ian M. Winter[1]

[1]*Centre for the Neural Basis of Hearing, The Physiological Laboratory, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK*

[2] *Laboratoire des Systèmes Perceptifs CNRS UMR 8248, École normale supérieure, Paris Sciences et Lettres Research University, Paris, France.*

*Corresponding author: np64@nyu.edu

ABSTRACT

Frequency modulation (FM) is a common acoustic feature of natural sounds, known to play a role in robust sound-source recognition. Auditory neurons show precise stimulus-synchronized discharge patterns that may be used for the representation of low-rate FM. However, it remains unclear whether this representation is based on 1) synchronization to slow temporal-envelope (ENV) cues resulting from cochlear filtering, or 2) phase locking to faster temporal-fine-structure (TFS) cues. To investigate the plausibility of those encoding schemes, single-units of the ventral cochlear nucleus of guinea pigs of either sex were recorded in response to sine-FM tones centered at the unit's best frequency (BF). The results show that for modulation depths within the receptive field, low BF units (<4 kHz) demonstrate good phase locking to TFS in contrast to high BF units. For modulation depths extending beyond the receptive field, the discharge patterns follow the ENV and fluctuate at the modulation rate. The receptive field proved to be a good predictor of the ENV responses for primary-like and chopper units. The current *in vivo* data also reveal a high level of diversity in responses across unit types. TFS cues are mainly conveyed by low-frequency and primary-like units and ENV cues by chopper and onset units. The diversity of responses exhibited by cochlear nucleus neurons provides a neural basis for a dual-coding scheme of FM in the brainstem, based on both ENV and TFS cues.

Keywords: frequency modulation, phase locking, envelope, temporal-fine-structure, cochlear nucleus

SIGNIFICANCE STATEMENT

Natural sounds, including speech, convey informative temporal modulations in frequency. Understanding how the auditory system represents those frequency modulations (FM) has important implications, as robust sound-source recognition depends crucially on the reception of low-rate FM cues. Here, we recorded 115 single-unit responses from the ventral cochlear nucleus in response to FM and provide the first physiological evidence of a dual-coding mechanism of FM via synchronization to temporal-envelope cues, and phase locking to temporal-fine-structure cues. We also demonstrate a diversity of neural responses, with different coding specializations. These results support the dual-coding scheme proposed by psychophysicists to account for FM sensitivity in humans and provide new insights on how this might be implemented in the early stages of the auditory pathway.

INTRODUCTION

It is generally agreed that the auditory system is adapted and optimized for the encoding of naturalistic stimuli (Nelken et al., 1999; Lewicki, 2002; Woolley et al., 2005; McDermott and Simoncelli, 2011). Amongst the features characterizing natural sounds, low-rate frequency modulation (FM) may play a specific role. Consistent with this view, salient FM together with other forms of temporal modulations (e.g., amplitude modulation (AM)), are systematically found at low rates, <20 Hz, in speech and animal vocalizations, as well as in environmental and musical sounds (Attias and Schreiner, 1997; Wang, 2000; Singh and Theunissen, 2003; Rees and Malmierca, 2005; Varnet et al., 2017). Moreover, there is clear evidence that speech recognition performance in quiet and in the presence of background sounds are constrained by human auditory sensitivity to low-rate FM (e.g., Zeng et al., 2005; Binns and Cullin, 2007; Ruggles et al., 2011; Johannesen et al., 2016).

Numerous psychophysical studies have investigated the detection of low-rate sinusoidal frequency-modulation (SFM). Zwicker (1952; 1956) and Maiwald (1976a, 1976b) put forth an 'excitation-pattern model', whereby SFM is perceived via *temporal-envelope cues* (ENV). This mechanism is often referred to as 'FM-to-AM conversion', because frequency-dependent attenuation of the FM caused by the cochlear filters results in AM (Saberi and Hafter, 1995). However, the excitation-pattern model has often been challenged and several studies have demonstrated that changes over time in the pattern of neural phase locking to *temporal-fine-structure* (TFS) cues may be used to perceive SFM at low carrier frequencies (Demany and Semal, 1986, 1989; Moore and Sek, 1996; Whiteford and Oxenham, 2015; Paraouty et al., 2016; Paraouty and Lorenzi, 2017). This additional mechanism is assumed to be 'sluggish' and restricted to the processing of low-rate (<5-10 Hz) FM (Moore and Sek, 1996).

Neurophysiological studies have addressed this issue by using frequency sweeps, but knowledge regarding the underlying mechanisms of SFM coding in the early auditory pathway is relatively sparse. In addition, most studies examining FM responses in the cochlear nucleus (CN) predate the detailed physiological and morphological classifications of CN neurons. Responses of single auditory nerve fibres (ANFs) to FM sweeps have been studied in the cat (Britt and Starr, 1976; Sinex and Geisler, 1981) and were described as similar to responses to pure tones, i.e., ANFs discharged for each frequency transition that crossed the response area. At the CN level, Britt and Starr (1976) described the responses of primary-like units to FM sweeps as simple relays, while onset and pauser units responded more to one direction of sweep. Only a few studies by Møller (1972a, 1972b) examined the responses of CN units to SFM and showed that the response patterns were synchronized to the ENV. Fernald and Gerstein (1972) also showed that the mean discharge patterns of CN units followed the modulations of the ENV cues in response to triangular periodic FM. To our knowledge, no former study has examined and characterized neural phase locking to TFS in the responses of CN neurons to low-carrier and low-rate SFM stimuli.

This work aims at narrowing the gap between the psychophysical findings regarding SFM coding (e.g., Whiteford and Oxenham, 2015; Paraouty et al., 2016) and the physiological responses of auditory neurons to SFM. This was achieved by the characterization of the relative contributions of temporal-envelope (ENV) and temporal-fine-structure (TFS) coding of ventral cochlear nucleus (VCN) neurons, with a wide range of best frequencies (BF= 0.14 – 22 kHz) in response to low-rate SFM (<10 Hz). The results demonstrate the capacity of VCN neurons to encode FM information using both ENV and TFS cues. The data further show contrasting ENV

and TFS specializations in different unit types, providing a possible neural basis for a dual-encoding scheme of FM in the early auditory pathway.

MATERIALS AND METHOD

The preparation

Experiments were performed on 10 male and 17 female pigmented guinea pigs (Cavia porcellus), weighing between 300 and 800 g. The animals were anaesthetised with urethane (1.0 g/kg, ip) and hypnorm (or fentanyl) was administered as supplementary analgesia (1 ml/kg, im). Anaesthesia and analgesia were maintained at sufficient depth to abolish the pedal withdrawal reflex of the front paw. Additional doses of hypnorm or urethane were administered on indication. Core temperature was monitored with a rectal probe and maintained at 38 °C using a thermostatically controlled heating blanket (Harvard Apparatus). The trachea was cannulated and on signs of suppressed respiration, the animal was ventilated artificially with a pump (Bioscience, UK). Surgical preparation and recordings took place in a sound-attenuated chamber (Industrial Acoustics Company). The animal was placed in a stereotaxic frame, which had ear bars coupled to hollow speculae designed for the guinea pig ear. A mid-sagittal scalp incision was made and the periosteum and the muscles attached to the temporal and occipital bones were removed. The bone overlaying the left bulla was fenestrated and a silver-coated wire was inserted into the bulla to contact the round window of the cochlea for monitoring compound action potentials (CAP). The hole was resealed with Vaseline. The CAP threshold was determined at selected frequencies at the start of the experiment and thereafter upon indication. If the thresholds had deteriorated by more than 10 dB and were non-recoverable (for example, by removing fluid from the bulla), the experiment was terminated. A craniotomy was performed

exposing the left cerebellum. The overlying dura was removed and the exposed cerebellum was partially aspirated to reveal the underlying cochlear nucleus. The hole left from the aspiration was then filled with 1.5% agar in saline to prevent desiccation. The experiments performed in this study have been carried out under the terms and conditions of the project licence issued by the United Kingdom Home Office to the last author.

Neural recordings

Responses of single units were recorded extracellularly with glass-coated tungsten microelectrodes (Merrill and Ainsworth, 1972; Microelectrodes.net). Electrodes were advanced in the sagittal plane by a hydraulic microdrive (650 W; David Kopf Instruments) at an angle of 45°. Neural spikes were discriminated and stored as spike times and were analysed off-line using custom-written Matlab programs (The MathWorks). Single units were isolated using broadband noise as search stimulus. All stimuli were digitally synthesized in real-time with a PC equipped with a DIGI 9636 PCI card that was connected optically to an AD/DA converter (ADI-8 DS; RME audio products). The AD/DA converter was used for digital-to-analog conversion of the stimuli as well as for analog-to-digital conversion of the amplified (1000x) neural activity. The sample rate was 96 kHz. The AD/DA converter was driven using ASIO (Audio Streaming Input Output) and SDK (Software Developer Kit) from Steinberg (Lloyd, 2002). After digital-to-analog conversion, the stimuli were equalized (phonic graphic equalizer, model EQ 3600; Apple Sound) to compensate for the speaker and coupler frequency response and fed into a power amplifier (Rotel RB971) and a programmable end attenuator (0–75 dB in 5 dB steps, custom build) before being presented over a speaker (Radio Shack 30-1777 tweeter assembled by Mike Ravicz, Massachusetts Institute of Technology, Cambridge, MA) mounted in the coupler

designed for the ear of a guinea pig. The stimuli were monitored acoustically using a condenser microphone (Bruël & Kjær 4134) attached to a calibrated 1-mm diameter probe tube that was inserted into the speculum close to the eardrum.

Unit classification

Upon isolation of a unit, its best frequency (BF) and excitatory threshold were first determined manually using audio-visual criteria and were verified off-line using an automated fitting procedure. The receptive field (or response map) for each unit was computed from 50-ms responses to pure tones played for a set of different stimulation levels: from 14 to 94 dB SPL in 5-dB steps and for a set of frequencies below and above the unit's BF: 2 and 3 octaves, respectively for BF <5 kHz and 1 and 2 octaves, respectively when BF >5 kHz (in 0.1 steps per octave). Both level and frequency were randomly presented. Peri-stimulus time histograms (PSTHs) with a binwidth of 0.2 ms were generated from spike times collected in response to 250 sweeps of a 50-ms tone (with randomized starting phase and 1-ms raised-cosine ramps) at the unit's BF at 20 and 50 dB above threshold.,. The tone bursts were repeated with a period of 250 ms. Spontaneous activity was measured over a 10-second period. Single units were classified based on their pure-tone PSTHs, the first order interspike-interval (ISI) distribution and the coefficient of variation (CV) of the discharge regularity.

The CV was calculated by averaging the ratios of the standard deviation divided by the mean ISI between 12 and 20 ms after onset (Young et al., 1988; Wright et al., 2011). On the basis of differences in the CV, the population of chopper units was divided into sustained choppers (CS, CV <0.3), and transient choppers (CT, CV ≥0.3) (Blackburn and Sachs, 1989). All units were classified as primary-like (PL), primary-like with notch (PN), chopper-sustained (CS), chopper-

transient (CT), onset-chopper (OC) and other onset types (onset: O, onset-L: OL, and onset-I: OI). The onset units were sub-divided according to the scheme introduced by Winter and Palmer (1995). For some units with very low BFs (~<0.5 kHz), it was not possible to assign them to one of the above categories. In the absence of a definitive classification, these are grouped together as 'low-frequency' (LF) units. For the population data, all recorded units were categorized into three major groups: 1) *Primary-likes and Low-frequency* (including PL, PN and LF units), 2) *Choppers* (including CS and CT units), and 3) *Onsets* (including OC, O, OL, and OI units). In this study, no other types of units were included (for instance, pauser, buildup or other dorsal cochlear nucleus response patterns).

Sinusoidal frequency-modulated stimuli

A sinusoidal frequency modulation (SFM) was imposed on a pure-tone stimulus (the carrier), with frequency (*fc*) set at the unit's BF. The modulation rate (*fm*) was 2, 5 or 10 Hz. Modulation depth (Δf) was 2, 4, 8, 16, or 32% relative to the BF. SFM tones were 1-second long, including 5-ms raised-cosine ramps at the start and end of the stimuli. The time interval between two stimuli was 1 second and the presentation level was set to 55 dB SPL. The SFM was presented at positive and negative starting polarities (Φ_C: starting carrier phases), while the starting phase of the modulator (Φ_M) was fixed (see equation below and Fig. 1). For each carrier phase, responses to 25 presentations of SFM stimuli were recorded. All the different experimental conditions were randomized. SFM responses at different stimulation levels were also recorded when possible (number of units, N=29 for lower sound levels (20-45 dB SPL) and N=40 for higher sound levels (60-90 dB SPL), compared to 55 dB SPL). In addition, for a subset of 15 units, SFM tones were played off-BF, with *fc* above and/or below the BF of the unit (from 0.5 to 2 octaves).

SFM stimuli = $\sin [(2\pi * fc * t + \Phi_C) + \beta * \cos(2\pi * fm * t + \Phi_M)]$,

with $\beta = \Delta f / fm$,

Δf =2, 4, 8, 16, and 32% * fc,

fm= 2, 5, and 10 Hz,

Φ_C = 0 and 180°, and

Φ_M = 0°.

Please insert Fig. 1.

Analyses

Spike times collected in response to 25 sweeps of the 1-second SFM stimuli were analysed and SFM-PSTHs were generated for the two different starting polarities (0 and 180°). Period histograms to the modulation rate were computed for the 0°-starting-phase condition, as well as the vector strengths. To avoid onset effects, spikes in response to the first modulation cycle were discarded.

In order to examine the respective contributions of ENV and TFS coding for each unit, shuffled correlograms were computed (Joris, 2003; Joris et al., 2006). Shuffled correlograms provide a smoother representation of the temporal characteristics present in the neural responses compared to standard all-order interval histograms (Louage et al., 2004). To compute the *shuffled auto-correlograms* (SACs), spike trains to repeated presentations of the SFM stimulus were compared pairwise by counting the number of instances that spikes were fired at the same instant in time (i.e., coincidences). Starting with the first spike of the first spike train, all forward intervals between this reference spike and all other spikes in non-identical spike trains were

measured, and tallied in a histogram. Only intervals across spike trains were considered, while intervals within spike trains were excluded in order to avoid the obscuring effect of the refractory period. In counting the number of coincidences, a 50 micro-second window was defined over which two spikes were regarded as being coincident (Joris et al., 2006). The whole procedure was repeated for all spikes in all spike trains and again, all forward intervals between the reference spike and all other spikes in non-identical spike trains were measured, and tallied in the same histogram. SACs were then normalized (Joris et al., 2006; Louage et al., 2004) such that the bin values were independent of average firing rate r, number of presentations N, choice of bin width Δt, and stimulus duration D. This normalized number of coincidences was achieved by dividing by $N(N-1)r^2 \Delta t D$. Here, N corresponded to 25 and Δt to 0.00005 second and D was 1 second. The SAC is displayed symmetrical around 0 ms; each positive interval of spike train pair (sweep 1, sweep 2) has a negative interval in pair (i.e., sweep 2, sweep 1). A peak height of 1 of the normalized SAC at 0 ms delay indicates a lack of stimulus-induced temporal structure. Larger values indicate that the spike times tend to be correlated between the different spike trains, and lower values indicate anti-correlation (Joris et al., 2006).

Like SACs, *shuffled cross-correlograms* (XACs) are also similar to all-order interval histograms, but here, the spike times are compared across responses to two different stimuli: the standard stimulus with $\Phi_C = 0°$ and the polarity-inverted stimulus, with $\Phi_C = 180°$, rather than across responses to the same stimulus (as for the SAC, in which responses to only the standard stimulus are examined). XACs were normalized by $N^2 r_1 r_2 \Delta t D$, where r_1 corresponded to the mean firing rate to the presentation of the standard stimulus and r_2 corresponded to the mean firing rate to the presentation of the polarity-inverted stimulus. A peak height of 1 at 0 ms delay

of the normalized XAC indicates a lack of stimulus-induced temporal structure, similarly to the SAC.

The peak heights of the SAC and the XAC indicate the strength of temporal coding, of either ENV or TFS or any mixture of both ENV and TFS. In order to disambiguate and quantify the strength of TFS and ENV coding, the *Sumcor* and *Difcor* are computed (Joris et al., 2006; Heinz and Swaminathan, 2009). The Sumcor is the average of the SAC and the XAC, while the Difcor is the difference between SAC and XAC. The response component that changes upon inverting the polarity (i.e., Φ_C) is due to synchronization to TFS, whereas the response component common to the standard stimulus ($\Phi_C = 0°$) and the polarity-inverted stimulus ($\Phi_C = 180°$) reflects synchronization to ENV. By taking the average of the SAC and XAC (i.e., the Sumcor), the common contribution of ENV coding is emphasized and the contribution of TFS coding is minimized. For the Difcor, a value of 0 indicates the number of coincidences expected from chance (rather than a value of 1 as in the other correlograms: SAC, XAC and Sumcor). The TFS contributions do not always cancel out completely in the Sumcor (Heinz and Swaminathan, 2009). This leakage of TFS into the Sumcor reflects distortion that arises from rectification associated with neural responses. The undesirable contribution of TFS coding to the Sumcor was eliminated according to Heinz and Swaminathan (2009) by considering only the envelope spectra below CF. In addition, the ENV contributions do not always cancel out completely in the Difcor (Heinz and Swaminathan, 2009). However, the influence of ENV coding on the Difcor can be argued to be small based on the small effect of sound level on Difcor peak heights (see Heinz and Swaminathan, 2009; Louage et al., 2004). In contrast, ENV coding degrades significantly with increasing sound level (Joris and Yin 1992; Louage et al. 2004).

The shuffled correlogram analyses are applicable to any repeatable stimulus, for instance, AM tones (Kale and Heinz, 2010), as well as to broadband noise (Joris 2003; Louage et al. 2004; 2005; Swaminathan and Heinz, 2011), or chimaeric speech (Heinz and Swaminathan, 2009). They are also widely used to analyse responses to monaural stimuli (e.g., Joris, 2003; Heinz and Swaminathan, 2009; Kale and Heinz, 2010, Swaminathan and Heinz, 2012). A limitation of these stationary shuffled correlogram analyses is that they only estimate the overall strength of ENV and TFS averaged across the whole duration of the SFM stimuli (1000 ms). In other words, they do not account for the temporal dynamics of the SFM stimuli. Thus, a sliding short-time analysis was also developed to explore the 'non-stationarity' of the temporal structure of the SFM stimulus. *Shuffled-all-order-ISI histograms* were computed using a windowing procedure, similar to Sayles et al. (2015). The analysis window was centered over 50 ms, and positive inter-spike intervals (ISIs) were analysed in this 50-ms time bin, creating an ISI histogram. The latter was computed based on the calculation of ISIs between ordered pairs of non-identical spike trains. A sliding window of 5 ms was used and the 1000-ms response recorded was analysed. The normalized number of coincidences (normalization factor: $N(N-1)r^2 \Delta t\, D$) for each time-window analysed were computed. A *running correlogram* was then built from those normalized ISI histograms, showing the modulated ISI distribution as a function of time (i.e., SFM duration=1000 ms).

Statistical Methods

All statistical analyses were computed using STATISTICA software (StatSoft France, 2013). *T-tests* for independent samples were used for comparison between data sets in Fig. 5E. A p-value of <0.05 was used for the significance limit.

A one-factor analysis of variance (ANOVA) was conducted, with dependent variable corresponding to the Sumcor peak heights of all units analysed, in order to assess the differences between unit types, regarding the strength of ENV coding (Fig. 8A). The factor 'unit type' had 7 levels for the different unit types: LF, PL, PN, CS, CT, OC, and O (see Unit classification). Post-hoc tests with Bonferroni corrections were also computed. Similarly, another one-factor ANOVA was conducted with dependent variable corresponding to the Difcor peak heights of all analysed units, in order to assess the differences between unit types, regarding the strength of TFS coding (Fig. 8B).

A repeated-measures ANOVA was conducted on the Sumcor peak heights (and the Difcor peak heights), in order to assess the effect of the modulation rate of the SFM (Figs. 9A and 9B). The ANOVA was computed separately for the three major groups of units: 1) *Primary-likes and Low-frequency*, 2) *Choppers*, and 3) *Onsets*. For all three ANOVAs, the factor 'modulation rate' had 3 levels corresponding to 2, 5, and 10 Hz. Similarly, repeated-measures ANOVAs were also conducted on the Sumcor peak heights (and the Difcor peak heights) for the three major unit types separately, in order to assess the effect of the modulation depth of the SFM (Figs. 9C and 9D). The factor 'modulation depth' for the three ANOVAs had 5 levels corresponding to 2, 4, 8, 16 and 32 %.

Modeling

The purpose of the modeling was to assess to what extent the receptive field of a unit can predict the ENV responses observed (i.e., resulting from FM-to-AM conversion). The raw receptive field measured for each unit was thus used to predict the amount of ENV fluctuations in the PSTHs in response to SFM. For each SFM condition (i.e., the 3 modulation rates and 5

297 modulation depths), the firing rates were modeled from the receptive field and plotted on the
298 recorded PSTH (see Fig. 10A). In addition, for each SFM condition, the correlation between the
299 recorded PSTH and the predicted PSTH was computed, as well as an overall root-mean-square
300 error (rmse). The model was also used to predict the level dependence of ENV responses, as well
301 as off-BF responses (Fig. 10).

302 It is important to note that the model did not include other peripheral factors (e.g., short-term
303 neural adaptation, amplitude compression, or lateral suppression) and CN factors (e.g., intrinsic
304 neural properties, neural circuitry) which also contribute to the responses of CN units to SFM.
305 The effects of those factors can be indirectly observed from what this simple model cannot
306 predict.

307

308 **RESULTS**

309 115 neurons were recorded in the VCN in response to SFM stimulus, played at different
310 modulation rates (*fm* = 2, 5, and 10 Hz) and different modulation depths (Δf = 2, 4, 8, 16, and
311 32% of BF). The carrier frequency was adjusted for each unit and set to the unit's BF (i.e., *fc*=
312 BF). PSTHs and period histograms to the modulation rate were computed. In addition, the
313 responses of each unit were analysed using shuffled auto- and cross-correlograms (SAC and
314 XAC), and the strengths of ENV and TFS coding were quantified using the peak heights of the
315 Sumcor and Difcor, respectively. The relative strength of ENV to TFS was calculated using the
316 ratio of the XAC to SAC. Data from single units are first shown (Figs. 2-7), followed by the
317 population analyses (Figs. 8-9), and the modeled responses (Fig. 10). All units were classified as:
318 LF (number of units, N=16), PL (N=23), PN (N=19), CS (N=9), CT (N=23), OC (N=9), and
319 other onset-type (including O, OL and OI, N=6).

ENV synchronized responses

Please insert Fig. 2.

An example of a high-BF PL unit (13.3 kHz) is shown in Fig. 2, with the pure-tone PSTH presented at 20 dB above the unit's threshold (Fig. 2A), together with its receptive field (Fig. 2B). Fig. 2C shows the PSTHs in response to the 1-second long SFM tone played at 55 dB SPL at 5 modulation depths: 2, 4, 8, 16 and 32% of BF, and 3 modulation rates: 2, 5, and 10 Hz. At 2 and 4% modulation depths, the PSTHs show no obvious ENV-following response, however at 8% depth (frequency sweeping from 12.2 kHz to 14.3 kHz), the mean discharge rate starts to follow the ENV cues resulting from FM-to-AM conversion; i.e., the mean firing rate is modulated. At that presentation level (55 dB SPL), the low frequency edge of the response area corresponds to 11.5 kHz and the high frequency edge to 15.2 kHz. The frequencies swept by the SFM approach both the low and high edges of the receptive field when the modulation depth is between 8 to 16%. At 32% depth (frequency sweeping from 9.0 kHz to 17.5 kHz), the PSTH represents the SFM fully sweeping in and out of the receptive field, as the firing rate drops to zero when the frequencies of the SFM are outside the receptive field. Fig. 2D illustrates schematically how the period histograms to the modulation rate in Fig. 2E are constructed. The top plot shows the period histogram computed from the raw data obtained, while the bottom plot shows the period histogram with the first half cycle representing '*sweep-down*', i.e., when the instantaneous frequency goes from high to low frequencies (also referred to as 'downward-going'), and the second half cycle representing '*sweep-up*', i.e., when the instantaneous frequency goes from low to high frequencies (also referred to as 'upward-going'). In Fig. 2E, all period histograms are computed as the latter plot. The period histograms to the modulation rate are shown for the various modulation rates and depths as in Fig. 2C, after excluding the response

to the first cycle of the stimuli. To assess the strength of synchronization to the ENV, the vector strength was calculated (Huffman et al., 1998). Significant vector strengths (i.e., $p<0.05$ according to Rayleigh's criterion) are indicated by an asterisk. The period was taken (2*1/f) as the ENV had twice the modulation frequency of the stimulus. The ENV responses were more salient for the higher modulation rates and depths. There was no direction preference and the firing rate was similar when the SFM sweeps from high-to-low or from low-to-high frequencies. Fig. 2F shows the normalized shuffled correlogram (SAC and XAC) in response to a 5-Hz SFM at 32% depth. The SAC and XAC overlap completely. Fig. 2G shows the normalized Sumcors for the 32%-depth SFM, at 3 different modulation rates: 2, 5 and 10 Hz. The peak height of the Sumcor increases as the modulation rate increases (from *fm*=2 to 10 Hz), showing better ENV coding at the highest modulation rate. The normalized Sumcors for the 5-Hz modulation rate SFM, at the 5 different modulation depths: 2, 4, 8, 16, 32% are shown in Fig. 2H. The ENV representation, as assessed by the Sumcor peak height, is higher at 32% depth, compared to the lower modulation depths. Overall, SFM (at the highest modulation rates and depths) can be represented via synchronization to the ENV cues in CN units, as previously described (in the cat- Fernald and Gerstein, 1972; in the rat- Møller, 1972a, 1972b). The Difcor (not shown) is flat for this 13.3-kHz PL unit, showing no TFS coding at all, as the difference between the SAC and XAC (Fig. 2F) leads to 0 in this case.

Phase locking to TFS cues

Please insert Fig. 3.

An example of responses of a LF unit (BF= 372 Hz) is given in Fig. 3. The PSTHs in response to the SFM (Fig. 3C) showed no obvious ENV-following response. This is expected as

the frequencies swept by the SFM stimuli remain within the receptive field, even at 32% modulation depth. At this modulation depth, the SFM stimuli sweeps from 253 Hz to 491 Hz, well within the low (141 Hz) and high (797 Hz) frequency edges of the receptive field. Fig. 3E shows the normalized shuffled correlogram (SAC and XAC) in response to a 5-Hz SFM at 32% depth and Fig. 3F and 3G represent the normalized Difcors (SAC–XAC) for the different modulation depths and rates. The damped oscillatory shaped Difcor has the same frequency as the BF of the unit, reflecting the carrier frequency of the SFM. The unit is phase locking to the carrier of the stimulus (i.e., to the TFS). The Difcor peak heights are similar for all modulation rates and depths, indicating similar strengths of TFS coding. The Sumcor in this case (not shown) is flat and has a value of 1, indicating no ENV coding. Fig. 3D shows the running correlograms (or interval histograms) computed in order to account for the temporal dynamics of the SFM stimuli (see Methods). The running correlograms are shown for the same experimental conditions as in Fig. 3C (i.e., 3 modulation rates: 2, 5, 10 Hz and 3 modulation depths: 2, 16, 32%). The TFS information is represented in the temporal dynamics of the firing pattern of the LF unit. The first peak (corresponding to the smallest ISI) in all the running correlograms occurs at 2.7 ms, which is equivalent to the *fc* of the SFM (also equivalent to the unit's BF= 372 Hz). The changes in instantaneous frequency of the stimuli with time are well represented in the spike timings, with 2, 5, and 10 cycles for the 2, 5, and 10-Hz SFM respectively in 1 second. The modulation depths, i.e., the frequencies swept by the stimuli, are also well represented in the running correlograms. Overall, the TFS information is present in the ISI of low-BF units and is conveyed to higher structures of the auditory brainstem.

Level dependence of ENV and TFS coding

Please insert Fig. 4.

The responses as a function of sound level are shown in Fig. 4C for a CS unit (BF= 9.5 kHz), at 3 levels over a 40 dB range. Like the PL unit shown in Fig. 2, at low modulation depths (2%), the PSTHs are flat and an ENV-following response emerges from the 4% depth condition. This unit also illustrates a common finding amongst the units recorded, the phenomenon of 'peak separation' in the PSTH (i.e., the doublets of the peaks in the PSTH). This is consistent with data recorded from cat CN (Fernald and Gerstein, 1972) and reflects the asymmetry of the receptive field, particularly at high levels. As the SFM tone sweeps in and out of the response area, at high levels, most of the energy is outside at the high frequency end (firing rate decreases to zero), while still remaining within at the low frequency (only a small decrease in firing rate). This phenomenon is well predicted by the model developed in the current study, based on the receptive field as shown in Fig. 10B. In Fig. 4D, the Sumcor peak height at zero increases with decreasing stimulus level reflecting better ENV-following responses at lower levels. The bandwidth of the receptive field is sharper at low levels compared to higher levels, hence the ENV representation is sharper at low levels. Another example of a high BF unit (PL unit with BF= 9.2 kHz, Fig. 4E) is given in response to SFM played at 5-Hz rate and 32% deviation. For this PL unit as well, the peak height of the Sumcor increases with decreasing stimulation level, showing better ENV representation at low levels. When considering the population data for all recorded units, there is a significant difference in Sumcor peak height (paired *t-test*: p=0.009) between responses evoked at lower sound levels (<55 dB SPL) and responses evoked at higher sound levels (>55 dB SPL).

The Difcors of a LF unit (BF= 417 Hz) in response to the SFM played at different levels are shown in Fig. 4F. The Difcor peak heights are similar at all three levels. When considering the

population data, there is no significant difference in Difcor peak heights for all recorded units (*t-test*: p=0.429) between responses evoked at lower sound levels (<55 dB SPL) and responses evoked at higher sound levels (>55 dB SPL). The TFS-based representation is hence very similar at different stimulation levels above threshold (Johnson, 1980). This is consistent with Palmer and Russell (1986), and their data from ANFs of the guinea pig, whereby phase locking, as measured by vector strength, increased with stimulation level and reached a saturation point about 20 dB above threshold. In addition, the Difcor peak heights were found to be maximal at around 50-70 dB SPL (Louage et al., 2004) using broadband noise. There was a slight decrease in TFS coding at higher sound levels, attributed to the effects of (cochlear) amplitude compression (Javel et al., 1983; Greenwood 1986).

Asymmetric responses of Onset units

Please insert Fig. 5.

PL and PN units, as well as CT and CS units discharge similarly for both the upward- and downward-going parts of the SFM. However, this was often not the case for units classified as *Onsets* (e.g., Winter and Palmer, 1995). An example of an OC unit with an asymmetric ENV-following response is given in Fig. 5. The first half cycle of the period histograms to the modulation rate (D) corresponds to the responses for the downward-going part of the SFM, i.e., from high to low frequencies; while the second half cycle corresponds to the responses for the upward-going part of the SFM, i.e., from low to high frequencies. The unit discharges preferentially to the downward-going part of the SFM stimulus (i.e., in the same direction as the green arrow on the receptive field). A direction selectivity index (DSI) was calculated as (number of spikes for sweep-up) – (number of spikes for sweep-down) divided by the (total

number of spikes for sweeps-up and down), as described by Mendelson and Cynader (1985). For the current OC, the DSI was -0.17, for a 10 Hz-SFM rate and 32% depth.

Fig. 5E shows the DSI for all units in response to the SFM played at 10 Hz merged across all modulation depths. The units are separated in 3 main groups: 1) open circles representing Primary-likes and Low-frequency units (i.e., PL, PN and LF), 2) grey triangles representing Chopper units (i.e., CS and CT), and 3) squares representing Onset units (with black squares for OC, red squares for OL, and blue squares for OI and O). The asymmetric responses of the present Onset units altogether are quite small (DSI values <0.3) in comparison to the asymmetric responses obtained in the inferior colliculus (IC) of the bat (DSI values >0.6), in response to SFM (e.g., Casseday et al., 1997) and the auditory cortex of the cat (>0.3), using upward- and downward-going FM sweeps at different speeds (e.g., Mendelson and Cynader, 1985). Nevertheless, the values of DSI obtained for the current Onset units are significantly different to those of Primary-likes and Low-frequency units (t-test for independent samples: $p<0.0001$) and Chopper units ($p<0.0001$).

It is important to point out that some OI units did not 'respond' to the SFM, except with one precise first spike at the beginning of the stimulus and very few spikes afterwards. Fig. 6A shows the pure-tone PSTH presented at 50 dB above the unit's threshold, followed by the receptive field (Fig. 6B) of an OI unit. Fig. 6C shows the PSTHs in response to the SFM played at 55 dB SPL, at various modulation rates and depths. Very few spikes are obtained in most conditions and an ongoing response can only be seen at the highest modulation depths (16 and 32%).

Please insert Fig. 6.

Off-BF responses

Please insert Fig. 7.

For a subset of 15 units, responses to off-BF SFM were also recorded. An example is given in Fig. 7 for a CS unit. The PSTHs are plotted for different modulation rates and depths on-BF (Fig. 7B, *fc*=5.4 kHz) and off-BF, with a carrier frequency at approximately half an octave below BF (Fig. 7C, *fc*=3.8 kHz), in order for the energy of the SFM stimulus to fall within the tail of the receptive field. The PSTH shape changes with off-BF stimulation and only one peak is present in the PSTH instead of two peaks (as in the on-BF condition). In other words, off-BF responses show modulations in their PSTH at the SFM rate, whereas on-BF responses are modulated at twice the SFM rate. The shape of the off-BF responses can be accurately predicted from the receptive field of the unit (see Fig. 10D). In addition, the ENV following response in the PSTH is present at a lower modulation depth condition (8%), whereas for the on-BF stimuli, the PSTHs are modulated only at 16% modulation depth condition. There is a trend for enhanced ENV-following responses (i.e., fluctuations in the PSTH) in off-BF conditions at low modulation depths, which are also predicted from the receptive field. However, at the 32% modulation depth, the peak heights of the Sumcors were very similar (data not shown) across the two conditions (on- and off-BF presentations). The Difcors also remained flat in both cases, showing no more TFS processing in the tail of the receptive field compared to the tip (data not shown). This was consistent across several units examined, and may be due to the low sound levels used (55 dB SPL). Altogether, these data are partly consistent with the notion, initially formulated by psychophysicists, that FM detection can be achieved by monitoring off-frequency channels tuned to lower (or higher) frequencies than the carrier frequency (Zwicker, 1956; Ernst & Moore, 2010).

Different coding specializations for different unit types

Please insert Fig. 8.

The peak heights of the Sumcors (Fig. 8A) and the Difcors (Fig. 8B) at zero delay are shown in Fig. 8 for all units. The Sumcor peak heights correspond to the strength of the ENV-based representation in the neural response and the Difcor peak heights correspond to the strength of the TFS-based representation in the neural response. Different symbols indicate the three main unit types: 1) *Primary-likes and Low-frequency units* (PL, PN and LF), 2) *Chopper units* (CS and CT), and 3) *Onset units* (OC, O, OL, and OI) in response to SFM at a modulation rate of 10 Hz and a modulation depth of 32% (stimulation level= 55 dB SPL).

Different unit types have different ENV and TFS-following responses. Onset and Chopper units are the best ENV encoders, while PL and PN units, and particularly, LF units are the best TFS encoders. Two one-factor ANOVAs revealed a significant effect of 'unit type', both for the ENV- and the TFS-following responses (ENV-response: [$F(6,150)= 23.6$, $p<0.0001$], TFS-response: [$F(6,150)= 30.8$, $p<0.0001$]). Onset units are significantly better than all other unit types for ENV coding (confirmed by post-hoc tests with Bonferroni correction and by *t-test*, e.g., when comparing O and CT, $p<0.0001$), while Chopper units are significantly better than Primary-like units (e.g., comparing CT and PN, $p= 0.008$), which are in turn significantly better than LF units (comparing PL and LF, $p<0.0001$). LF units are significantly better than all other unit types for TFS coding (e.g., comparing LF and PL, $p<0.0001$), while Primary-like units are significantly better than Chopper units (e.g., comparing PL and CT, $p=0.023$), which are relatively similar to Onset units (e.g., comparing CT and O, $p=0.638$).

Ratio of ENV and TFS coding

The shuffled correlograms (SAC and XAC) allow quantification of the individual strengths of ENV and TFS coding by the Sumcor and Difcor metrics and the relative strength of ENV and TFS coding can be quantified by the ratio of the peak heights of XAC to SAC at zero delay (see Louage et al. 2004; 2005; Kale and Heinz, 2010). A ratio of 0 represents primarily TFS coding, while a ratio of 1 represents primarily ENV coding. For a low BF unit, the SAC and the XAC are inverted in polarity (see Fig. 3E), leading to an XAC/SAC ratio close to 0, whereas for a high BF unit, the SAC and XAC are overlapping (see Fig. 2E) leading to an XAC/SAC ratio close to 1.

The XAC/SAC ratio for all units is shown in Fig. 8C as a function of BF, in response to SFM played at 55 dB SPL, with modulation rate= 5 Hz and modulation depth= 2%. At this modulation depth, all of the energy of the stimulus remained within the receptive field of the unit, even for high BF units. The transition region whereby units change from a more TFS-based response to a more ENV-based response ranges from 1 to 4 kHz. TFS coding is no longer present beyond 4 kHz. Similarly to Louage et al. (2005), responses are defined as ENV-dominated for XAC/SAC ratio $\geq$0.9, and the frequency cut-off value at this ratio corresponds to ~4 kHz for Primary-like and Low-frequency units. Chopper and Onset units show higher ENV-following responses in comparison to Primary-like units, at least for BFs <2 kHz. Those units are poor TFS-encoders, but good ENV-encoders. Thus, they have a lower transition region from TFS to ENV coding compared to Primary-like and Low-frequency units.

Since the SFM used in the current study was at fixed modulation depths and not adjusted to the bandwidth of the unit under study, the transition from TFS to ENV can be thought of to be mainly due to the bandwidth of the unit. In other words, the sharper the bandwidth, the more salient ENV cues, resulting from FM-to-AM conversion are present, independently of unit types. Fig. 8D shows the distribution of the XAC/SAC ratio as a function of Q10 (or Q_{10dB}) calculated

from the receptive fields of all units. Indeed, for Primary-like and Low-frequency units, the transition in ratio observed is quite similar to the transition observed in Fig. 8C. In contrast, for Onset and Chopper units, even when the SFM stimuli are well within the filter bandwidth, those units do not encode the TFS information as well as the Primary-like and Low-frequency units. In other words, at similar Q10 values as Primary-like and Low-frequency units, the XAC/SAC ratio for Chopper and Onset units are higher compared to those of Primary-like and Low-frequency units. There is, however, quite a large variability in the Q10 values. Nevertheless, some Onset and Chopper units, despite having relatively small Q10 values (<3) have an XAC/SAC ratio of 1. Altogether, this suggests that the filter bandwidth (expressed here as Q10) does indeed constrain FM-to-AM conversion and the strength of ENV coding. However, Chopper and Onset units show enhanced ENV coding, and reduced TFS coding at similar Q10 or BF values as Primary-like and Low-frequency units, demonstrating coding specializations as well.

FM coding as a function of modulation rate and depth

Please insert Fig. 9.

The Sumcor (Fig. 9A) and the Difcor (Fig. 9B) peak heights at zero delay for the different unit types as a function of the modulation rate (2, 5, and 10 Hz), in response to SFM at a fixed modulation depth of 32% are shown in Fig. 9. The ENV-based representation and the TFS-based representation are similar across modulation rates, for all unit categories. A repeated-measures ANOVA revealed no significant effect of modulation rate for Onset units ([$F(2,22)= 2.7$, $p=0.09$]).

Fig. 9 also shows the Sumcor (Fig. 9C) and Difcor (Fig. 9D) peak heights for the different unit types as a function of the modulation depth (2 to 32%), in response to SFM at a fixed

modulation rate of 5 Hz. While the TFS-based representation is constant across modulation depths, the ENV-based representation is significantly enhanced with modulation depth for all three unit categories (repeated-measures ANOVA: for Primary-like and Low-frequency units: [$F(4,296)= 30.7$, $p<0.0001$], for Chopper units: [$F(4,220)= 66.8$, $p<0.0001$], and for Onset units: [$F(4,52)= 4.7$, $p=0.003$]). Significant differences are obtained between the responses of particular unit types at different modulation depths (for instance, Primary-like and Low-frequency units from 2 to 16%, paired *t-test*: $p= 0.010$, Chopper units from 2 to 8%, $p= 0.001$). At higher modulation depths, the SFM stimuli sweep across a wider range frequencies and hence, the possibility of crossing the edges of the response area of a particular unit are increased, leading to more ENV cues.

Modeling results

Please insert Fig. 10.

In order to quantify the ENV representation at different modulation rates and depths, the raw receptive field (i.e., without any smoothing) of each unit was used to predict the FM-to-AM conversion, in other words, the ENV fluctuations in the PSTH. Fig. 10A shows the PSTH of the PL unit from Fig. 2 and the predicted ENV responses modeled from the receptive field of the PL unit. Differences in overall spiking rate are expected since the receptive field is computed from 50-ms responses to pure tones, while the PSTH is computed from 1000-ms responses to SFM. In addition, the model did not take into account any physiological characteristic of CN neurons nor physiological peripheral processes. However, the shape of the PSTHs and the ENV responses (the fluctuations in the PSTHs) are well predicted as quantified by the correlation value (e.g., $p<0.001$ for 2-, 5- and 10-Hz SFM conditions at 32% depth). The overall root-mean-square error

(rmse) for this unit is equal to 51.1. The differences in spiking rate between the data and the predictions cause the rmse to be rather large, despite the fact that the ENV shape is quite accurately modeled. At 2 and 4% modulation depths, the predicted PSTHs are flat, showing no obvious ENV-following response, consistent with the original data (except for the onset response and the adaptation observed). At higher modulation depths (starting at 8%), the PSTH shows fluctuations, well modeled from the receptive field.

Level differences in ENV coding can also be well predicted from the receptive field, as shown in Fig. 10B. The PSTHs of the CS unit from Fig. 4 are shown together with the predicted ENV responses at 75, 55, and 35 dB SPL. The filter shape and bandwidth can predict the sharper ENV responses at decreasing stimulation level ($p<0.001$ for 32% depth condition at all 3 levels). Fig. 10C shows the PSTHs of the OC unit from Fig. 5, with an asymmetric response to the upward- and downward-going parts of the SFM. The measurements derived from the receptive field do not predict this asymmetry. Off-BF responses, on the other hand, can be accurately predicted as shown in Fig. 10D. The overall firing rate is not accurate, however the shape of the predicted PSTHs match closely those from the original data of the CS unit from Fig. 7C. Overall, a simple model based on the receptive field can accurately predict the shape of the ENV fluctuations in the PSTH for all conditions presented here, except for the asymmetric ENV responses.

DISCUSSION

The present study assessed the relative contributions of ENV and TFS-coding for a population of VCN single units in response to low-rate SFM. The results provide clear evidence that FM can be encoded via: 1) synchronization to ENV cues, generated at the output of cochlear

596 filters and represented in the fluctuations of the mean firing pattern, and 2) phase locking to TFS
597 cues and represented in the precise spike timing. The diversity in the responses of different unit
598 types provides new insights regarding how this dual-coding scheme might be implemented in the
599 early auditory system.
600
601 **Unit specialization for ENV and TFS coding**
602 The data show that Onset units (multipolar and octopus cells) are specialized in ENV coding.
603 At similar BFs, Onset units show higher ENV synchronization than Chopper units, which in turn
604 show higher ENV responses than Primary-like units (Onset > Chopper > Primary-like). This is
605 consistent with previous studies demonstrating the hierarchy of ENV representation in the CN
606 using AM stimuli (Frisina et al., 1990; Rhode and Greenberg, 1994; Wang and Sachs, 1994;
607 Joris et al., 2004). Precise inhibitory circuits have been proposed to underlie ENV enhancement
608 at various levels of the auditory pathway (Koch and Grothe, 1998; Backoff et al., 1999; Krishna
609 and Semple, 2000; Caspary et al., 2002; Ter-Mikaelian et al., 2007). Onset-I units (octopus cells)
610 have been shown to be exceptional AM encoders (Rhode, 1994; Golding et al., 1995). Here,
611 most onset-I units only fired at the onset of the stimuli. This lack of response may be due to the
612 low modulation rates and depths used here, as octopus cells are particularly sensitive to the rate
613 of depolarization (Ferragamo and Oertel, 2002) and have been shown to fire when a wide array
614 of ANFs are synchronously active (Oertel, 2005).
615 Low-frequency and Primary-like units (bushy cells) with BFs <3 kHz show strong phase
616 locking to TFS cues, consistent with the notion that bushy cells provide fast-fluctuating TFS
617 information to the superior olivary complex and form part of the binaural sound-localization
618 stream (Yin, 2002). The current data also show that the strength of phase locking to TFS was

relatively independent of the modulation rate, depth and level of stimulation of the SFM. Hence, compared to ENV cues which are highly dependent on modulation depth (Fig. 7C) and stimulation level (Fig. 3), TFS cues provide an invariant and robust code. However, F0-related periodicity information have been shown to be degraded in the presence of both reverberation and F0 modulation (Sayles and Winter, 2008; Sayles et al., 2015).

The significant degradation of ENV coding with increasing stimulation level is in line with ANF data (Joris and Yin, 1992; Wang and Sachs, 1994; Louage et al., 2004; Dreyer and Delgutte, 2006), and is consistent with the saturating character of rate-level functions. In the current study, the observed level responses are due to the combined effect of saturation and receptive field bandwidth at different stimulation levels. In other words, the strength of ENV elicited from FM-to-AM conversion decreases as the stimulus level increases since the tuning of CN neurons are broader at higher sound levels compared to lower ones.

Limit of phase locking to TFS

A coding scheme for FM based on TFS cues would be restricted to relatively low carriers due to the limit of neural phase-locking. In the current data, the transition region where units' responses change from being TFS-driven to being ENV-driven is ~4 kHz (Fig. 6, XAC/SAC ratio= 0.9) for Primary-like and Low-frequency units. The upper limit of phase-locking in ANFs, assessed with vector strength measurements, is known to vary across species; 5-6 kHz in squirrel monkeys and cats (Rose et al., 1967; Johnson, 1980) and 4-5 kHz in guinea pigs and chinchillas (Harrison and Evans, 1979; Palmer and Russell, 1986). The owl is exceptional in this respect, since phase locking is constant up to 9-10 kHz (Sullivan and Konishi, 1984; Köppl, 1997).

In agreement with previous reports examining temporal coding for a population of ANFs (Louage et al., 2004; Kale and Heinz, 2010), the current CN data show a sigmoidal relationship between the ratio of TFS- to ENV-coding (i.e., the XAC/SAC ratio) and frequency (i.e., BF). The transition frequency from a TFS-based to an ENV-based representation found here is consistent with ANF responses to sinusoidal-AM in the chinchilla (~3 kHz, Kale and Heinz, 2010). In response to broadband noise, ANF responses showed the same trend (Louage et al., 2004), with a cutoff of ~5 kHz in the cat. Louage et al. (2005) examined the responses of trapezoid-body fibres and reported a lower cutoff (~4 kHz) for Primary-like responses compared to ANFs. Indeed, several studies have shown a decrease in the phase-locking cutoff along the ascending auditory pathway (Nelson et al., 1966; Schuller, 1979; Rees and Møller, 1983; Gaese and Ostwald, 1995; Lu and Wang, 2000). This decrease has been proposed to reflect the conversion of temporally-synchronized cues into a rate-based representation, allowing the integration of auditory information with other sensory input at the cortical level (Wang et al., 2008).

The human phase-locking cutoff is still unknown (at least ~3 kHz, Joris and Verschooten, 2013). In addition, SFM coding via synchronization to ENV cues is highly dependent on the frequency selectivity of the cochlear filters, as it constrains the amount of ENV. It is very likely that the current results underestimate ENV cues from FM-to-AM conversion in humans, since the latter have been estimated to have sharper tuning (2-3 times) than cats, chinchillas and guinea pigs (Shera et al., 2002; Oxenham and Shera, 2003; Joris et al., 2011; Manley and van Dijk, 2016; however see Ruggero and Temchin, 2005, 2007). Hence, the dual-coding scheme of FM demonstrated here in the guinea pig should be similar or even more efficient for humans, at least regarding ENV coding when considering the output of several overlapping sharp auditory filters.

From the current model predictions (Fig. 10), the global fluctuations in spiking rate (ENV responses) are well accounted for when considering only FM-to-AM conversion. However, the small differences between the data and the predictions, as well as the absence of asymmetric responses from the current model suggest that additional mechanisms (e.g., intrinsic neural properties of CN units, neural circuitry of the VCN) and more complex models of CN units (e.g., Manis and Campagnola, 2018) need to be considered.

Direction selectivity

Physiological data regarding direction selectivity in the auditory system appear to be dependent on species, as well as on the recording site. In contrast with the visual system where direction selectivity is observed at the very beginning of sensory processing (Fried et al., 2002), ANFs show symmetrical discharge patterns to ascending and descending parts of FM signals, suggesting a lack of direction preference (Britt and Starr, 1976; Sinex and Geisler, 1981). However, as early as the CN, neurons show asymmetrical responses (Erulkar et al., 1968; Møller, 1974a; 1974b; Godfrey et al., 1975; Britt and Starr, 1976). These early reports did not unequivocally identify the class of unit type, or the region of the cochlear nucleus they were recorded from, making comparisons with the present study difficult. Consistent with some of these early reports, however, the current results show a small, but significant preference, for Onset units to respond preferentially to the descending part of the SFM (from high to low frequencies). Although direction selectivity was not prominent in the rat CN (Møller, 1969) a preference for descending sweeps was found at high sweep rates (Møller, 1971) and in the cat CN, a small preference for ascending sweeps was found in Onset units (Rhode and Smith, 1986). At the cortical level, neurons selective to ascending and descending directions are equally

687 abundant (Poon et al., 1991; Nelken and Versnel, 2000; Zhang et al., 2003; Kuo and Wu, 2012).
688 Auditory neurons of bats present a strong downward preference and it has been noted that
689 downward frequency sweeps are common in their echolocation calls (Suga, 1965; Razak and
690 Fuzessery, 2006; Andoni et al., 2007). Interestingly, some differences have also been reported in
691 the perception of rising and falling frequency sweeps for human listeners, with a preference for
692 rising glides (Collins and Cullen, 1984; Carlyon and Stubbs, 1989).
693 Previous studies (e.g., Suga, 1988) have suggested that the distribution of excitatory and
694 inhibitory regions may underlie direction selectivity, and higher DSI values may be obtained for
695 larger FM depths than examined here, as the effect of any inhibitory sideband would be
696 exacerbated. To our knowledge, there is no evidence for inhibition playing a role in the
697 responses of Onset units.
698 In a previous study by Smith and Rhode (1986), the OI unit category showed the strongest
699 direction selectivity. The arrangement of ANF inputs on the dendrites of octopus cells
700 (McGinley et al., 2012; Oertel et al., 2000) suggests that these cells may perform across-
701 frequency processing and this may contribute to the asymmetry. It is unclear why we failed to
702 observe this but it might be due to different stimulation parameters or species differences. In a
703 computational model of octopus cells (presumed OI units), Levy and Kipke (1997) also failed to
704 observe any directional selectivity. It is worth noting that in this paper, the onset classification
705 scheme adopted was that of Winter and Palmer (1995). From the Winter and Palmer data set, OC
706 and OL units were modelled as a continuum (e.g., Kalluri and Delgutte, 2003). Others, however,
707 have suggested that OL units are associated with globular bushy cells and form a continuum with
708 the PN response type (Spirou et al, 2005). Across-frequency coincidence detection has
709 previously been shown in PN units (Carney, 1990; Wang and Delgutte, 2012). They are thus

sensitive to local changes in the spatio-temporal pattern of AN activity. Monaural cross-frequency coincidence detection results in a temporal sharpening across frequency channels, which may be useful in pitch perception, although the importance of across-fiber spike timing remains unclear in monaural processing (however, see Carney ,1990; Heinz et al., 2001; Joris et al., 1994). It is possible that these properties would lead to a preference for sweep direction in PN units or the OL units hypothesized to form a continuum with them. The DSI values of the OL units in this paper are closer to the OC units than the PN units. Finally, other neural mechanisms such as adaptation (Bleeck et al., 2006; Ingham et al., 2016) or rate of depolarisation (McGinley and Oertel, 2006) could contribute to the observed directional selectivity.

A dual-code subserving FM detection

A better understanding of the underlying mechanisms of FM coding should improve our knowledge of how the auditory system processes speech and other natural sounds. The current results are consistent with data obtained for SFM detection in human listeners. For low carrier frequencies (<5 kHz) and modulation rates (<5-10 Hz), several studies suggest that human listeners mostly use TFS cues for SFM detection (Moore and Sek, 1996; Paraouty et al., 2016; Wallaert et al., 2016, 2017; Paraouty and Lorenzi, 2017). In comparison, for high carriers and rates, listeners seem to mostly use ENV cues (Ernst and Moore, 2010, 2012).

The current findings support the view that low-rate FM is encoded by two sensory mechanisms based on: 1) synchronization to slow ENV cues resulting from cochlear filtering, and 2) phase locking to faster TFS cues. The absence of differences in the TFS-following responses across modulation rates tested here (2-10 Hz) indicates that the locus of sluggishness for TFS processing postulated by psychophysicists (Moore and Sek, 1996) is more central than

733 the CN. A dual-coding scheme has several advantages compared to a single ENV coding
734 scheme. For high sound (conversational) levels, most low threshold, high spontaneous rate fibres
735 are saturated when the SFM stimuli is played at BF, leading to reduced ENV synchronization
736 (Sachs and Young, 1979; Joris and Yin, 1992). Off-BF responses, on the other hand, are less
737 affected at high sound levels (Wang and Sachs, 1993) and the ENV cues would still be salient.
738 Phase locking to TFS cues also remains relatively robust at high sound levels. In addition, in the
739 presence of noise or competing backgrounds, phase locking to TFS cues provides a more robust
740 representation in comparison to synchronized ENV responses (Shamma and Lorenzi, 2013). In
741 reverberant environments, TFS coding of linear frequency swept harmonic complexes was found
742 to be degraded, which in turn, impairs stream segregation (Winter and Sayles, 2008; Sayles et al.,
743 2015). Overall, phase locking to TFS cues is likely to play a crucial role in the robust
744 representation of speech and other ecologically important sounds in a broad range of acoustic
745 situations. The dual-coding scheme is thus adapted to the constraints of natural listening
746 conditions, which are constantly changing.
747

ACKNOWLEDGMENTS

N Paraouty and IM Winter were supported by a grant from Entendre SAS. N Paraouty, L Varnet and C Lorenzi were supported by ANR HEART, ANR-11-0001-02 PSL and ANR-10-LABX-0087. We thank A Stasiak (Microelectrodes.net) for providing electrodes for part of the recordings, and CG Ouanounou for help in analysis and figure editing.

REFERENCES

Andoni S, Li N, Pollak GD (2007) Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. J Neurosci 27:4882–4893.

Attias H, Schreiner CE (1997) Temporal Low-Order Statistics of Natural Sounds. Adv Neural Inf Process Syst 9:27–33.

Backoff PM, Palombi PS, Caspary DM (1999) GABA and glycinergic inputs shape coding of AM in chinchilla cochlear nucleus. Hear Res 134:77–88.

Binns C, Culling JF (2007) The role of fundamental frequency contours in the perception of speech against interfering speech. J Acoust Soc Am 122:1765–1776.

Blackburn CC, Sachs MB (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. J Neurophysiol 62:1303-1329.

Bleeck S, Sayles M, Ingham NJ, Winter IM (2006) The time course of recovery from suppression and facilitation from single units in the mammalian cochlear nucleus. Hear Res 212:176–184.

Britt RH, Starr A (1976) Synaptic events and discharge patterns of cochlear nucleus cells. II. Frequency-modulated tones. J Neurophysiol 39:179–194.

Carlyon RP, Stubbs RJ (1989) Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers. J Acoust Soc Am 85:2563–2574.

Carney LH. Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. J Neurophysiol 64: 437–456, 1990.

Caspary DM, Palombi PS, Hughes LF (2002) GABAergic inputs shape responses to amplitude modulated stimuli in the inferior colliculus. Hear Res 168:163–173.

776 Casseday JH, Covey E, Grothe B (1997) Neural selectivity and tuning for sinusoidal frequency
777 modulations in the inferior colliculus of the big brown bat, Eptesicus fuscus. J Neurophysiol
778 77:1595–1605.
779 Collins MJ, Cullen JKJ (1978) Temporal integration of tone glides. J Acoust Soc Am 63:469–
780 473.
781 Demany L, Semal C (1986) On the detection of amplitude modulation and frequency modulation
782 at low modulation frequencies. Acustica 61:243–255.
783 Demany L, Semal C (1989) Detection thresholds for sinusoidal frequency modulation. J Acoust
784 Soc Am 85:1295–1301.
785 Dreyer A, Delgutte B (2006) Phase locking of auditory-nerve fibers to the envelopes of high-
786 frequency sounds: implications for sound localization. J Neurophysiol 96:2327–2341.
787 Ernst SMA, Moore BCJ (2010) Mechanisms underlying the detection of frequency modulation. J
788 Acoust Soc Am 128:3642–3648.
789 Ernst SMA, Moore BCJ (2012) The role of time and place cues in the detection of frequency
790 modulation by hearing-impaired listeners. J Acoust Soc Am 131:4722–4731.
791 Erulkar SD, Butler RA, Gerstein GL (1968) Excitation and inhibition in cochlear nucleus. II.
792 Frequency-modulated tones. J Neurophysiol 31:537–548.
793 Fernald RD, Gerstein GL (1972) Response of cat cochlear nucleus neurons to frequency and
794 amplitude modulated tones. Brain Res 45:417–435.
795 Ferragamo MJ, Oertel D (2002) Octopus cells of the mammalian ventral cochlear nucleus sense
796 the rate of depolarization. J Neurophysiol 87:2262–2270.
797 Fried SI, Munch TA, Werblin FS (2002) Mechanisms and circuitry underlying directional
798 selectivity in the retina. Nature 420:411-414.

799 Frisina RD, Smith RL, Chamberlain SC (1990) Encoding of amplitude modulation in the gerbil
800 cochlear nucleus: I. A hierarchy of enhancement. Hear Res 44:99–122.
801 Gaese BH, Ostwald J (1995) Temporal coding of amplitude and frequency modulation in the rat
802 auditory cortex. Eur J Neurosci 7:438–450.
803 Godfrey DA, Kiang NYS, Norris BE (1975) Single unit activity in the posteroventral cochlear
804 nucleus of the cat. J Comp Neurol 162:247–268.
805 Golding NL, Robertson D, Oertel D (1995) Recordings from slices indicate that octopus cells of
806 the cochlear nucleus detect coincident firing of auditory nerve fibers with temporal precision. J
807 Neurosci 15:3138–3153.
808 Greenwood DD (1986) What is “Synchrony suppression”? J Acoust Soc Am 79:1857–1872
809 Harrison RV, Evans EF (1979) Some aspects of temporal coding by single cochlear fibers from
810 regions of cochlear hair cell degeneration in the guinea pig. Arch Otorhinolaryngol 224:71–78.
811 Heinz MG, Zhang X, Bruce IC, Carney LH (2001) Auditory nerve model for predicting
812 performance limits of normal and impaired listeners. Acoust Res Lett Online 2:91–96.
813 Heinz MG, Swaminathan J (2009) Quantifying Envelope and Fine-Structure Coding in Auditory
814 Nerve Responses to Chimaeric Speech. J Assoc Res Otolaryngol 10:407–423.
815 Huffman RF, Argeles PC, Covey E (1998) Processing of sinusoidally frequency modulated
816 signals in the nuclei of the lateral lemniscus of the big brown bat, Eptesicus fuscus. Hear Res
817 126:161–180.
818 Ingham NJ, Itatani N, Bleeck S, Winter IM (2016) Enhancement of forward suppression begins
819 in the ventral cochlear nucleus. Brain Res 1639:13–27.

Javel E, McGee J, Walsh EJ, Farley GR, Gorga MP (1983) Suppression of auditory nerve responses. II. Suppression threshold and growth, iso-suppression contours. J Acoust Soc Am 74:801–813.

Johannesen PT, Perez-Gonzalez P, Kalluri S, Blanco JL, Lopez-Poveda EA (2016) The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners. Trends Hear 20:1–14.

Johnson DH (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J Acoust Soc Am 68:1115–1122.

Joris PX, Yin TCT (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. J Acoust Soc Am 91:215.

Joris PX, Smith PH, Yin TCT (1994) Enhancement of neural synchronization in the anteroventral cochlear nucleus. II. Responses in the tuning curve tail. J Neurophysiol 71:1037–1051.

Joris PX (2003) Interaural time sensitivity dominated by cochlea-induced envelope patterns. J Neurosci 23: 6345-6350.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577.

Joris PX, Louage DH, Cardoen L, van der Heijden M (2006) Correlation index: a new metric to quantify temporal coding. Hear Res 216, 19-30.

Joris PX, Bergevin C, Kalluri R, Mc Laughlin M, Michelet P, van der Heijden M, Shera CA (2011) Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. Proc Natl Acad Sci 108:17516–17520.

Joris PX, Verschooten E (2013) On the limits of neural phase locking to fine structure in humans. Adv Exp Med Biol 787:101–108.

Kale S, Heinz MG (2010) Envelope coding in auditory nerve fibers following noise-induced hearing loss. J Assoc Res Otolaryngol 11:657–673.

Kalluri S, Delgutte B (2003) Mathematical models of cochlear nucleus onset neurons: II. Model with dynamic spike-blocking state. J Comput Neurosci 14:91–110.

Koch U, Grothe B (1998) GABAergic and glycinergic inhibition sharpens tuning for frequency modulations in the inferior colliculus of the big brown bat. J Neurophysiol 80(1):71–82.

Köppl C (1997) Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. J Neurosci 17:3312–3321.

Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. J Neurophysiol 84:255–273.

Kuo RI, Wu GK (2012) The Generation of Direction Selectivity in the Auditory System. Neuron 73:1016–1027.

Levy KL, Kipke DR (1997) A computational model of the cochlear nucleus octopus cell. J. Acoust. Soc. Am. 102, 391-402.

Lewicki MS (2002) Efficient coding of natural sounds. Nat Neurosci 5:356–363.

Louage DHG, van der Heijden M, Joris PX (2004) Temporal Properties of Responses to Broadband Noise in the Auditory Nerve. J Neurophysiol 91:2051–2065.

Louage DHG, van der Heijden M, Joris PX (2005) Enhanced Temporal Response Properties of Anteroventral Cochlear Nucleus Neurons to Broadband Noise. J Neurosci 25:1560–1570.

Lu T, Wang X (2000) Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat. J Neurophysiol 84:236–246.

865 Maiwald D (1967a) Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit
866 kleiner Frequenz und Amplitudenänderungen. Acustica 18:81-92.
867 Maiwald D (1967b) Die Berechnung von Modulationsschwellen mit Hilfe eines
868 Funktionsschemas. Acustica 18:193-207.
869 Manley GA, van Dijk P (2016) Frequency selectivity of the human cochlea: suppression tuning
870 of spontaneous otoacoustic emissions. Hear Res 336:53–62.
871 Manis PB, Campagnola L (2018) A biophysical modelling platform of the cochlear nucleus and
872 other auditory circuits: from channels to networks. Hear Res 360:76–91.
873 McDermott JH, Simoncelli EP (2011) Sound texture perception via statistics of the auditory
874 periphery: Evidence from sound synthesis. Neuron 71:926–940.
875 McGinley MJ, Oertel, D (2006) Rate thresholds determine the precision of temporal integration
876 in principal cells of the ventral cochlear nucleus. Hear Res. 216-217, 52-63.
877 McGinley MJ, Liberman MC, Bal R, Oertel D (2012) Generating synchrony from the
878 asynchronous: compensation for cochlear traveling wave delays by the dendrites of individual
879 brainstem neurons. J Neurosci 32:9301–9311.
880 Mendelson JR, Cynader MS (1985) Sensitivity of cat primary auditory cortex (AI) neurons to the
881 direction and rate of frequency modulation. Brain Res 327:331–335.
882 Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrodes. Med
883 Biol Eng 10:662–672.
884 Møller AR (1969) Unit Responses in the Cochlear Nucleus of the Rat to Sweep Tones. Acta
885 Physiol Scand 76:503–512.
886 Møller AR (1971) Unit Responses in the Rat Cochlear Nucleus to Tones of Rapidly Varying
887 Frequency and Amplitude. Acta physiol scand 81:540–556.

888 Møller AR (1972a) Coding of Amplitude and Frequency Modulated Sounds in the Cochlear
889 Nucleus of the Rat. Acta Physiol Scand 86:223–238.
890 Møller AR (1972b) Coding of sounds in lower levels of the auditory system. QRevBiophysics
891 5:59–155.
892 Møller AR (1974a) Coding of amplitude and frequency modulated sounds in the cochlear
893 nucleus. Acta Physiol Scand 31:292–299.
894 Møller AR (1974b) Coding of sounds with rapidly varying spectrum in the cochlear nucleus. J
895 Acoust Soc Am 55:631–640.
896 Moore BCJ, Sek A (1996) Detection of frequency modulation at low modulation rates: evidence
897 for a mechanism based on phase locking. J Acoust Soc Am 100:2320–2331.
898 Nelken I, Rotman Y, Bar Yosef O (1999) Responses of auditory-cortex neurons to structural
899 features of natural sounds. Nature 397:154–157.
900 Nelken I, Versnel H (2000) Responses to linear and logarithmic frequency-modulated sweeps in
901 ferret primary auditory cortex. Eur J Neurosci 12:549–562.
902 Nelson PG, Erulkar SD, Bryan JS (1966) Responses of units of the inferior colliculus to time-
903 varying acoustic stimuli. J Neurophysiol 29:834–860.
904 Oertel D, Bal R, Gardner SM, Smith PH, Joris PX (2000) Detection of synchrony in the activity
905 of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. Proc Natl Acad Sci
906 USA 97:11773–11779.
907 Oertel D (2005) Importance of timing for understanding speech. Focus on “perceptual
908 consequences of disrupted auditory nerve activity”. J Neurophysiol 93:3044–3045.
909 Oxenham AJ, Shera CA (2003) Estimates of Human Cochlear Tuning at Low Levels Using
910 Forward and Simultaneous Masking. JARO - J Assoc Res Otolaryngol 4:541–554.

911 Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its
912 relation to the receptor potential of inner hair-cells. Hear Res 24:1–15.
913 Paraouty N, Ewert SD, Wallaert N, Lorenzi C (2016) Interactions between amplitude modulation
914 and frequency modulation processing: Effects of age and hearing loss. J Acoust Soc Am
915 140:121–131.
916 Paraouty N, Lorenzi C (2017) Using individual differences to assess modulation-processing
917 mechanisms and age effects. Hear Res 344:38–49.
918 Poon P, Chen X, Hwang J (1991) Basic determinants of FM responses in the inferior colliculus
919 of rats. Exp Brain Res 83:598–606.
920 Razak KA, Fuzessery ZM (2006) Facilitatory Mechanisms Underlying Selectivity for the
921 Direction and Rate of Frequency Modulated Sweeps in the Auditory Cortex. J Neurophysiol
922 28:1303–1319.
923 Rees A, Malmierca MS (2005) Processing of Dynamic Spectral Properties of Sounds. Int Rev
924 Neurobiol 70:299–330.
925 Rees A, Møller AR (1983) Response of neurons in the inferior colliculus of the rat to AM and
926 FM tones. Hear Res 10:301–330.
927 Rhode WS (1994) Temporal coding of 200% amplitude modulated signals in the ventral cochlear
928 nucleus of cat. Hear Res 77:43–68.
929 Rhode WS, Greenberg S (1994) Encoding of amplitude modulation in the cochlear nucleus of
930 the cat. J Neurophysiol 71:1797–1825.
931 Rhode WS, Smith PH (1986) Encoding timing and intensity in the ventral cochlear nucleus of
932 the cat. J Neurophysiol 56:261–286.

933 Rose J, Brugge J, Anderson D, Hind J (1967) Phase-locked response to low-frequency tones in
934 single auditory nerve fibers of the squirrel monkey. J Neurophysiol 30:769–793.
935 Ruggero MA, Temchin AN (2005) Unexceptional sharpness of frequency tuning in the human
936 cochlea. Proc Natl Acad Sci USA 102:18614–18619.
937 Ruggero MA, Temchin AN (2007) Similarity of traveling-wave delays in the hearing organs of
938 humans and other tetrapods. J Assoc Res Otolaryngol 8:153–166.
939 Ruggles DR, Bharadwaj H, Shinn-Cunningham BG (2011) Normal hearing is not enough to
940 guarantee robust encoding of suprathreshold features important in everyday communication.
941 Proc Natl Acad Sci USA 108:15516–15521.
942 Saberi K, Hafter ER (1995) A common neural code for frequency and amplitude-modulated
943 sounds. Nature 374:537.
944 Sachs MB, Young ED (1979) Encoding of steady-state vowels in the auditory nerve:
945 Representation in terms of discharge rate. J Acoust Soc Am 66:470.
946 Sayles M, Stasiak A, Winter IM (2015) Reverberation impairs brainstem temporal
947 representations of voiced vowel sounds: challenging “periodicity-tagged” segregation of
948 competing speech in rooms. Front Syst Neurosci 8:1–19.
949 Schuller G (1979) Coding of small sinusoidal frequency and amplitude modulations in the
950 inferior colliculus of “CF-FM” bat, Rhinolophus ferrumequinum. Exp Brain Res 34:117–132.
951 Shamma SA, Lorenzi C (2013) On the balance of envelope and temporal fine structure in the
952 encoding of speech in the early auditory system. J Acoust Soc Am 133:2818–2833.
953 Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from
954 otoacoustic and behavioral measurements. Proc Natl Acad Sci USA 99:3318–3323.

955 Sinex DG, Geisler CD (1981) Auditory-nerve fiber responses to frequency-modulated tones.
956 Hear Res 4:127–148.
957 Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories
958 of auditory processing. J Acoust Soc Am 114:3394–3411.
959 Spirou GA, Rager J, Manis PB (2005) Convergence of auditory-nerve fiber projections onto
960 globular bushy cells. Neuroscience 136:843–863.
961 Suga N (1965) Analysis of Frequency-modulated sounds by auditory neurons of echo-locating
962 bats. J Physiol 179:26–53.
963 Suga N (1988). Auditory neuroethology and speech processing: Complex-sound processing by
964 combination-sensitive neurons. In: Auditory function: Neurobiological bases of hearing
965 (Edelman GM, Gall WE, Cowan WM), pp679-720. Eds. Wiley, New York.
966 Sullivan WE, Konishi M (1984) Segregation of stimulus phase and intensity coding in the
967 cochlear nucleus of the barn owl. J Neurosci 4:1787–1799.
968 Swaminathan J, Heinz MG (2011) Predicted effects of sensorineural hearing loss on across-fiber
969 envelope coding in the auditory nerve. J Acoust Soc Am 129:4001–4013.
970 Swaminathan J, Heinz MG (2012) Psychophysiological analyses demonstrate the importance of
971 neural envelope coding for speech perception in noise. J Neurosci 32:1747–1756.
972 Ter-Mikaelian M, Sanes DH, Semple MN (2007) Transformation of temporal properties between
973 auditory midbrain and cortex in the awake Mongolian gerbil. J Neurosci 27:6091–6102.
974 Varnet L, Ortiz-Barajas MC, Guevara Erra R, Gervain J, Lorenzi C (2017) A cross-linguistic
975 study of speech modulation spectra. J Acoust Soc Am 141(5): 3701-3702.
976 Wallaert N, Moore BCJ, Lorenzi C (2016) Comparing the effects of age on amplitude
977 modulation and frequency modulation detection. J Acoust Soc Am 139:3088–3096.

978 Wallaert N, Moore BCJ, Ewert SD, Lorenzi C (2017) Sensorineural hearing loss enhances
979 auditory sensitivity and temporal integration for amplitude modulation. J Acoust Soc Am
980 141:971–980.
981 Wang, X., Sachs, M. B. (1994). Neural encoding of single-formant stimuli in the cat. II.
982 Responses of anteroventral cochlear nucleus units. J Neurophysiol 71:59-78.
983 Wang, X., & Sachs, M. B. (1993). Neural encoding of single-formant stimuli in the cat. I.
984 Responses of auditory nerve fibers. J Neurophysiol 70:1054-1075.
985 Wang X (2000) On cortical coding of vocal communication sounds in primates. Proc Natl Acad
986 Sci USA 97:11843–11849.
987 Wang X, Lu T, Bendor D, Bartlett E (2008) Neural coding of temporal information in auditory
988 thalamus and cortex. Neuroscience 157:484–493.
989 Wang GI, Delgutte B (2012) Sensitivity of cochlear nucleus neurons to spatio-temporal changes
990 in auditory nerve activity. J Neurophysiol 108:3172–3195.
991 Whiteford KL, Oxenham AJ (2015) Using individual differences to test the role of temporal and
992 place cues in coding frequency modulation. J Acoust Soc Am 138:3093–3104.
993 Winter IM, Palmer AR (1995) Level dependence of cochlear nucleus onset unit responses and
994 facilitation by second tones or broadband noise. J Neurophysiol 73:141–159.
995 Woolley SMN, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal
996 modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci
997 8:1371–1379.
998 Wright MCM, Bleeck S, Winter IM (2011) An exact method of regularity analysis for auditory
999 brainstem neurons (L). J Acoust Soc Am 130:3545–3548.

1000 Young ED, Robert JM, Shofner WP (1988) Regularity and latency of units in ventral cochlear
1001 nucleus: implications for unit classification and generation of response properties. J
1002 Neurophysiol 60:1–29.
1003 Yin TC (2002) Neural mechanisms of encoding binaural localization cues in the auditory
1004 brainstem. In: Integrative functions in the mammalian auditory pathway, pp99-159. Springer
1005 New York.
1006 Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K (2005)
1007 Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci USA
1008 102:2293–2298.
1009 Zhang LI, Tan AYY, Schreiner CE, Merzenich MM (2003) Topography and synaptic shaping of
1010 direction selectivity in primary auditory cortex. Nature 424:201–205.
1011 Zwicker E (1952) Die Grenzen der Hörbarkeit der Amplitudenmodulation und der
1012 Frequenzmodulation eines Tones (The limits of audibility of amplitude modulation and
1013 frequency modulation of a pure tone). Acustica 2:125-133.
1014 Zwicker E (1956) Die elementaren Grundlagen zur Bestimmung der Informationskapazität des
1015 Gehörs (The elemental foundations for determining the information capacity of the auditory
1016 system). Acustica 6:365–381.
1017
1018

FIGURE CAPTIONS

Fig. 1.

Description of the sinusoidal frequency modulated (SFM) stimuli. ***A****, SFM, normalized in amplitude and plotted as a function of time from 25 to 75 ms. Black line shows the standard stimulus with* Φ_C *=0, and the red dotted line shows the polarity-inverted stimulus (i.e.,* Φ_C *= 180°).* ***B****, Instantaneous frequency of the SFM plotted in blue as a function of time from 0 to 1000 ms. Carrier frequency is 500 Hz (indicated with the red dotted line) and modulation rate (fm) is 10 Hz. The instantaneous frequency varies from 340 Hz to 660 Hz, as the modulation depth (Δf) is 32% of 500 Hz.*

Fig. 2.

Envelope responses (ENV) of a high BF unit characterized as PL, with best frequency (BF)= 13.3 kHz*, spontaneous rate (SR)=* 20.5 spikes/second*, and Threshold (T)=* 36 dB SPL. ***A****, Pure-tone PSTH (20 dB above threshold).* ***B****, Receptive field.* ***C****, PSTHs in response to SFM presented at 55 dB SPL at 5 modulations depths: 2, 4, 8, 16, and 32% (in columns) and 3 modulation rates: 2, 5, and 10 Hz (in rows), for fixed carrier at BF and modulator phase of 0°, and bindwidth= 0.2 ms.* ***D****, Schematic illustration of how the period histogram is computed, with the top plot showing raw data and the bottom plot after re-organization of the spike times in order to compare 'sweep-up' and 'sweep-down'. 'Sweep-up' correspond to an instantaneous frequency going upward and 'sweep-down' correspond to the instantaneous frequency going downward.* ***E****, Period histograms to the modulation rate for each of the different condition of modulation rates and depths (same as in C). Significant vector strength values are indicated by *, according to Rayleigh's criterion.* ***F****,* Normalized SAC (red) and XAC (black) *for 5-Hz SFM at modulation*

depth= 32%. **G**, *Sumcor (average SAC and XAC) for the 3 modulation rate conditions for a fixed modulation depth of 32% (2, 5, and 10 Hz in blue, red and black).* ***H****, Sumcor for the 5 modulation depth conditions for a fixed modulation rate of 5 Hz (2, 4, 8, 16, and 32% in grey, blue, green, red, and black).*

Fig. 3.

Temporal-fine-structure (TFS) responses of a low BF unit characterized as LF, with BF= 372 Hz, SR= 0.4 spikes/second, and T= 24 dB SPL. ***A****, Pure-tone PSTH (20 dB above threshold).* ***B****, Receptive field.* ***C****, PSTHs in response to 55-dB SPL SFM presented at 3 modulations depths: 2, 16, and 32% (in columns) and 3 modulation rates: 2, 5, and 10 Hz (in rows), for fixed carrier at BF and modulator phase of 0°, and bindwidth= 0.2 ms.* ***D****, Running correlograms for positive interspike intervals= 0 - 50 ms (see Methods). Responses to different rates and depths as illustrated in C. The color-scale bar applies to all 9 running correlograms in the figure.* ***E****, Normalized SAC (red) and XAC (black) for 5-Hz SFM at modulation depth= 32%.* ***F****, Difcor (SAC – XAC) corresponding to the 3 modulation rates conditions (2, 5, and 10 Hz) for a fixed modulation depth of 32%.* ***G****, Difcor corresponding to the 5 modulation depths conditions (2, 4, 8, 16, and 32%) for a fixed modulation rate of 5 Hz.*

Fig. 4.

Level dependence of the envelope responses (ENV) of a CS unit, with BF= 9.5 kHz, *SR=* 35.4 spikes/second, *T=* 23 dB SPL. ***A****, Pure-tone PSTH (20 dB above threshold).* ***B****, Receptive field.* ***C****, PSTHs in response to 5-Hz SFM presented at BF and at 3 stimulation levels: 35, 55, and 75 dB SPL (in rows) and 5 modulations depths: 2, 4, 8, 16, and 32% (in columns).* ***D****, Sumcor*

(average SAC and XAC) for 5-Hz SFM stimuli at modulation depth= 32%, at the 3 presentation levels (35, 55, and 75 dB SPL in red, blue and green). ***E****, Sumcor of another high-BF unit: primary-like (PL), BF= 9.2 kHz, at different presentation levels in dB SPL (in response to SFM stimuli: 5-Hz rate and 32% depth).* ***F****, Difcor of a low BF unit: low-frequency (LF), BF=417 Hz, at different presentation levels in dB SPL (in response to SFM stimuli: 5-Hz rate and 32% depth).*

Fig. 5.

Asymmetric envelope responses (ENV) of an OC unit, with BF=10.1 kHz, SR=0.9 spikes/second, T=47 dB SPL. ***A****, Pure-tone PSTH (50 dB above threshold).* ***B****, Receptive field, with a green arrow indicating direction of preference (for the downward-going part of the SFM).* ***C****, PSTHs in response to 55-dB SPL SFM presented at BF, at 3 modulations depths: 8, 16, and 32% (indicated in different columns) and 3 modulation rates: 2, 5, and 10 Hz (indicated in different rows).* ***D****, Period histograms to the modulation rate (similar experimental conditions as in C).* ***E****, Direction selectivity index for all units recorded (SFM at modulation rate= 10 Hz across all modulation depths). Three categories of units are described in the legend: 1) Primary-likes and Low-frequency (including PL, PN and LF units), in white circles, 2) Choppers (including CS and CT units), in grey triangles, and 3) Onsets (including OC, O, OL, and OI units), in squares (OC are represented in black, OL in red, and O and OI in blue). The BF of OC units varied between 4.1 kHz to 22.3 kHz, while those of OL units varied between 1.9 kHz to 14.7 kHz, and those of O and OI was 2.2 kHz and 9.1 kHz. Individual DSI values are plotted for all units, as well as box plots for the three unit categories. A DSI value of 0 corresponds to a symmetric response to both*

upward- and downward-going frequencies. A positive DSI value corresponds to a preference for upward-going sweep and a negative DSI value corresponds to a downward-going preference.

Fig. 6.

Some onset units respond primarily to the onset of the SFM. An OI unit is shown here, with BF= 9.1 kHz, SR= 0.0 spikes/second, T= 45 dB SPL. ***A****, Pure-tone PSTH (50 dB above threshold).* ***B****, Receptive field.* ***C****, PSTHs in response to 55-dB SPL SFM presented at BF, at 5 modulations depths: 2, 4, 8, 16, and 32% (in columns) and 3 modulation rates: 2, 5, and 10 Hz (in rows). An ongoing response can be seen only at the highest modulation depths (16 and 32%).*

Fig. 7.

Comparison of on- and off-BF responses of a CS unit, with BF=5.4 kHz, SR=71.0 spikes/second, T=20 dB SPL. ***A****, Receptive field. The solid line indicate the 'on-BF' position (carrier frequency at BF) and the dotted line the 'off-BF' position.* ***B****, PSTHs in response to 'on-BF' SFM presented at 55 dB SPL (horizontal dashed line) at 3 rates: 2, 5, and 10 Hz (in rows) and 5 modulations depths: 2, 4, 8, 16, and 32% (in columns). The carrier frequency (fc) was set to the same value as BF (5.4 kHz).* ***C****, 'PSTHs in response to 'off-BF SFM (same conditions as B); fc was set to 3.8 kHz.*

Fig. 8.

Population data: ENV- and TFS-following responses. ***A****, ENV-following response (peak height of Sumcor at zero delay) for all units. 7 unit types are shown: low-frequency (LF), primary-like (PL), primary-like with notch (PN), chopper-sustained (CS), chopper-transient (CT), onset-*

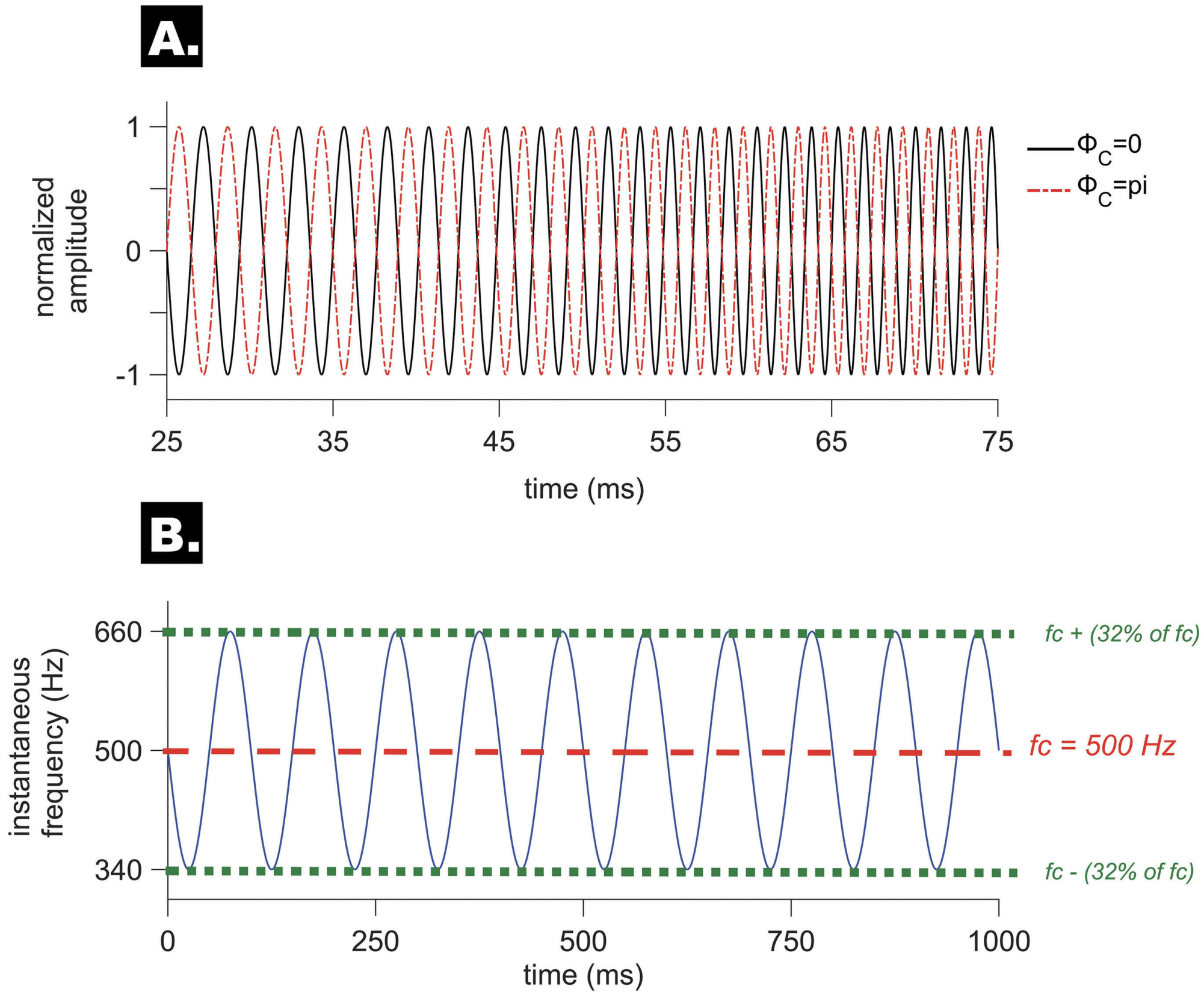
chopper (OC), and other onset types (including onset, onset-I and onset-L). Three main categories of units are described in the legend: 1) Primary-likes and Low-frequency (including PL, PN and LF units), in white circles, 2) Choppers (including CS and CT units), in grey triangles, and 3) Onsets (including OC, O, OL, and OI units), in black squares. Asterisks indicate significant differences between Sumcor values of different unit types. ***B****, TFS-following response (peak height of Difcor at zero delay) for all units. Same legend as for A.* ***C****, Ratio of XAC to SAC as a function of BF for all units recorded. A ratio of 0 indicates purely TFS-following response and a ratio of 1 indicates purely ENV-following response.* ***D****, Ratio of XAC to SAC as a function of Q10 (or Q_{10dB}) calculated from the receptive fields for all units recorded. Small Q10 values indicate sharp filter bandwidths and larger Q10 values indicate broader filter bandwidth.*

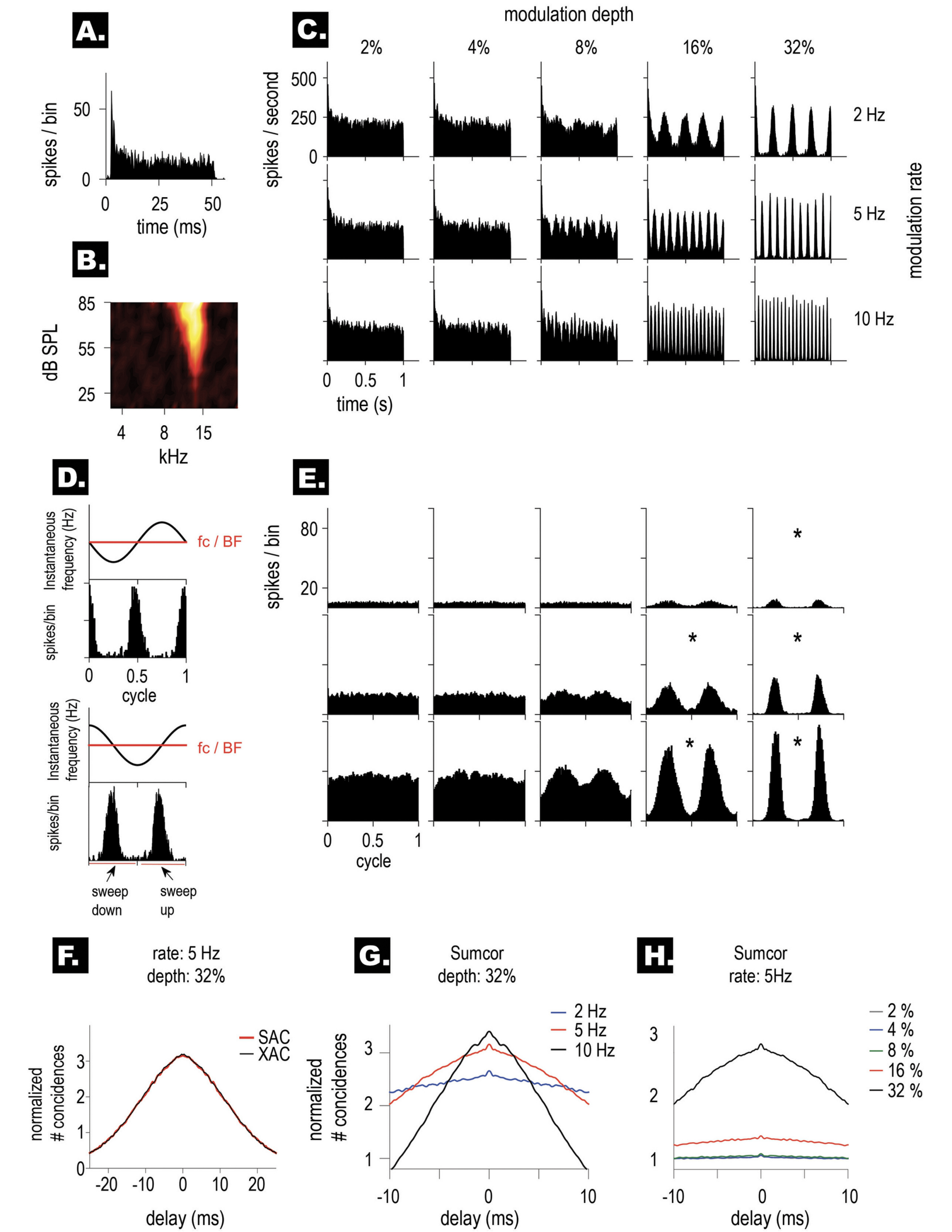
Fig. 9.

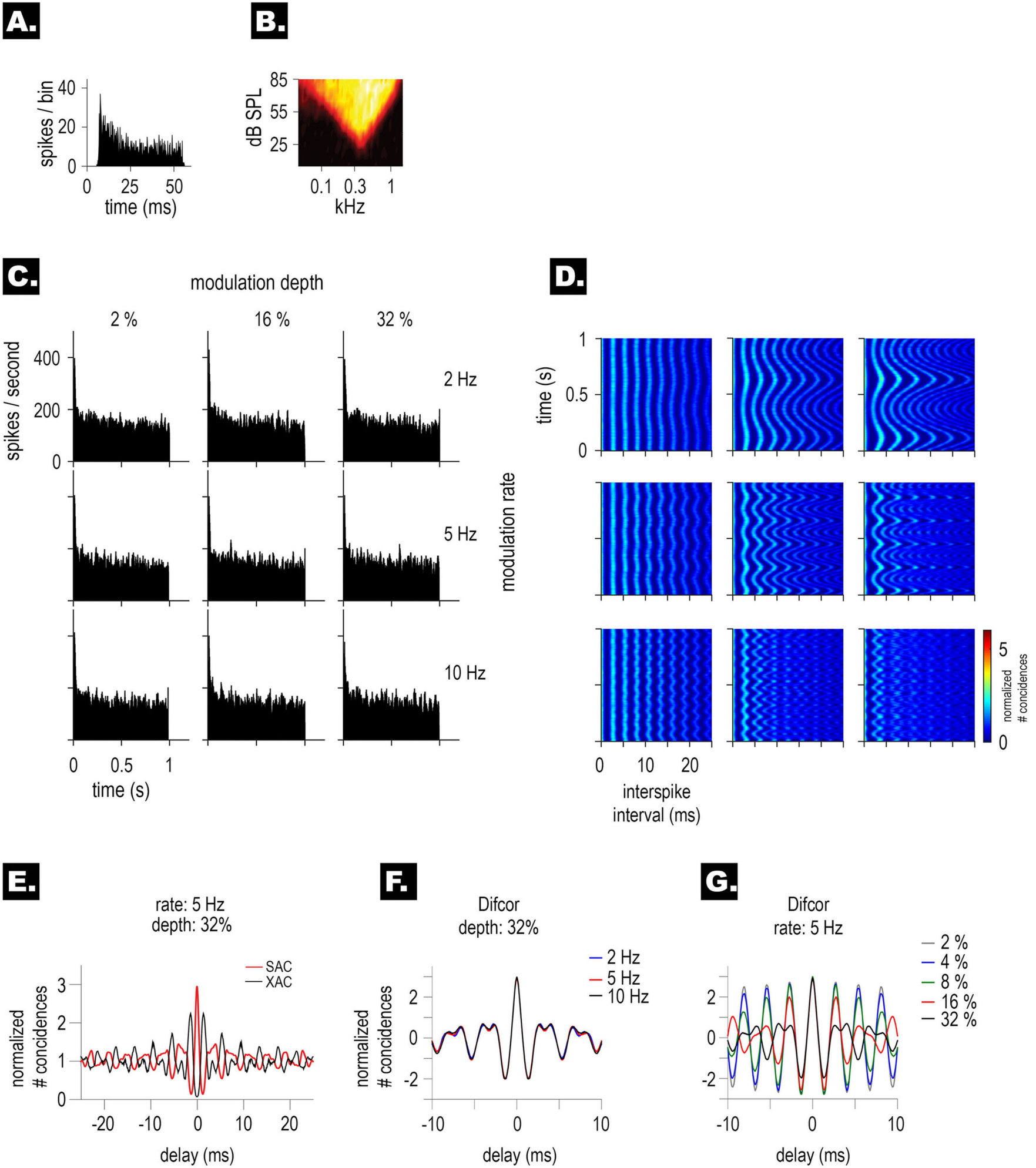
Strengths of ENV- and TFS-coding in response to SFM at different modulation rates and depths. ***A****, ENV-following response (peak height of the Sumcor at 0, see Methods) for each unit as a function of modulation rate: 2, 5, and 10 Hz, for a fixed modulation depth of 32%.* ***B****, TFS-following response (peak height of the Difcor at 0) for each unit as a function of modulation rate: 2, 5, and 10 Hz, for a fixed modulation depth of 32%.* ***C****, Envelope-following response for each unit as a function of modulation depth: 2, 4, 8, 16, and 32%, for a fixed modulation rate of 5 Hz.* ***D****, TFS-following response for each unit as a function of modulation depth: 2, 4, 8, 16, and 32%, for a fixed modulation rate of 5 Hz.*

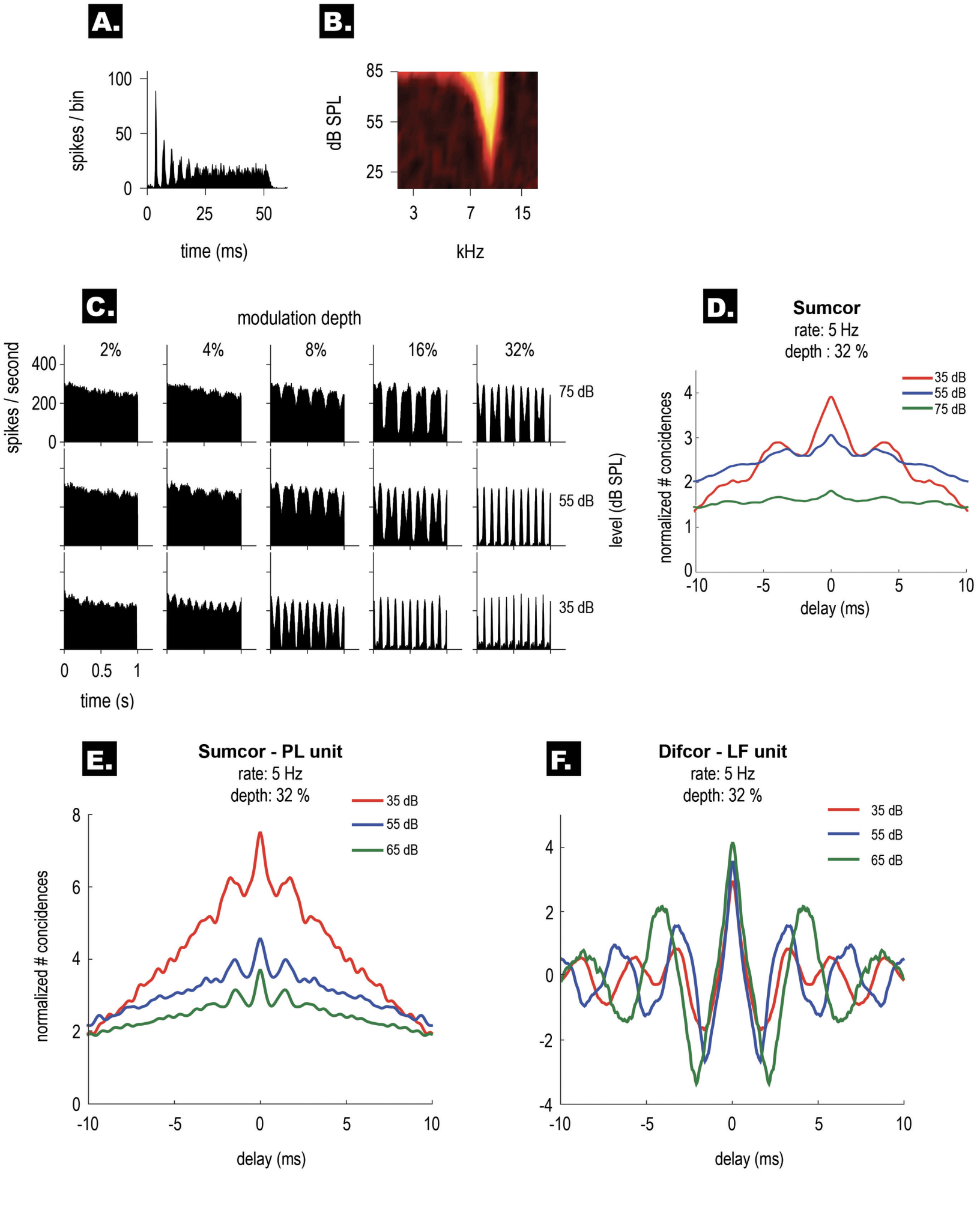
Fig. 10.

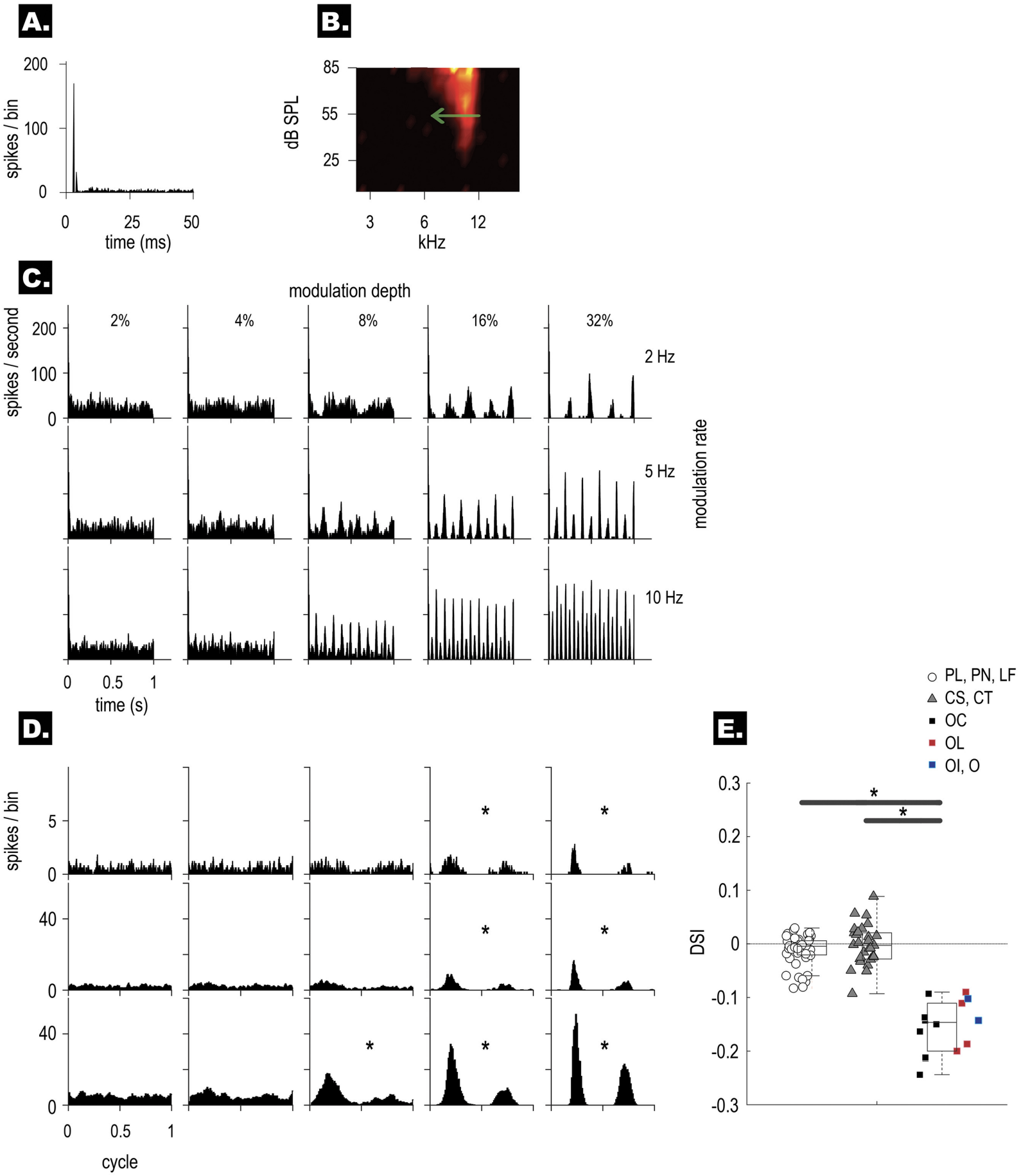
*Prediction of ENV responses from the raw receptive fields. **A**, PSTHs of the PL unit from Fig. 2 for different SFM conditions (black), with the predicted shape of the PSTHs from the receptive field of the PL unit (red). **B**, PSTHs of the CS unit from Fig. 4 for different SFM conditions (black) at 3 different levels (shown in the three rows), with the predicted shape of the PSTHs from the receptive field of the CS unit (red). **C**, PSTHs of the OC unit from Fig. 5 (black) showing asymmetric responses, with the predicted shape of the PSTHs from the receptive field of the OC unit (red). **D**, PSTHs of the CS unit from Fig. 7 for different SFM conditions (black) presented off-BF (Fig. 7C) and the predicted shape of the PSTHs from the receptive field of the CS unit (red).*

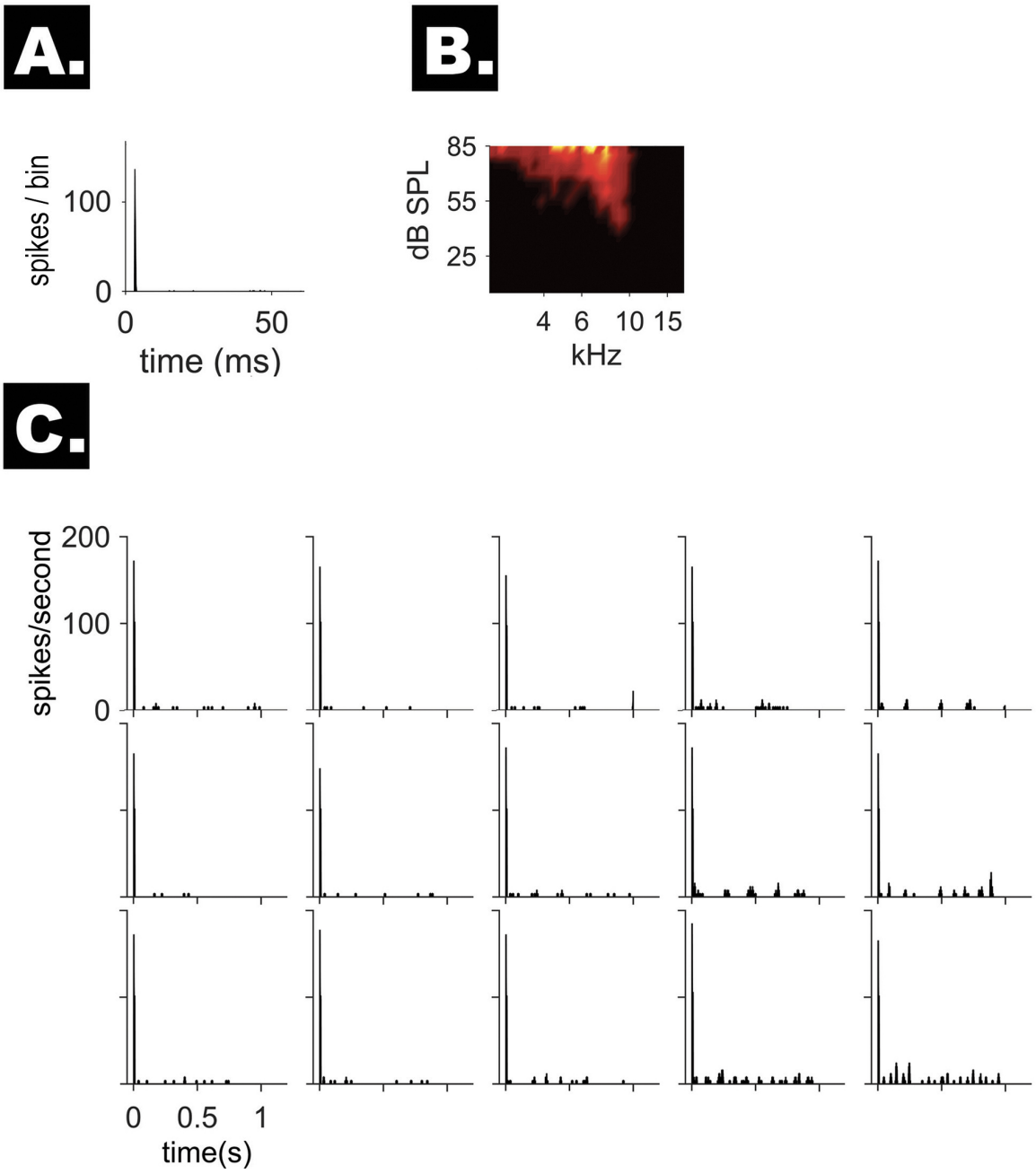

A.
1
0
-1
normalized amplitude
25
35
45
55
65
75
time (ms)
Φc=0
Φc=pi
B.
660
500
340
instantaneous frequency (Hz)
fc + (32% of fc)
fc = 500 Hz
fc - (32% of fc)
0
250
500
750
1000
time (ms)

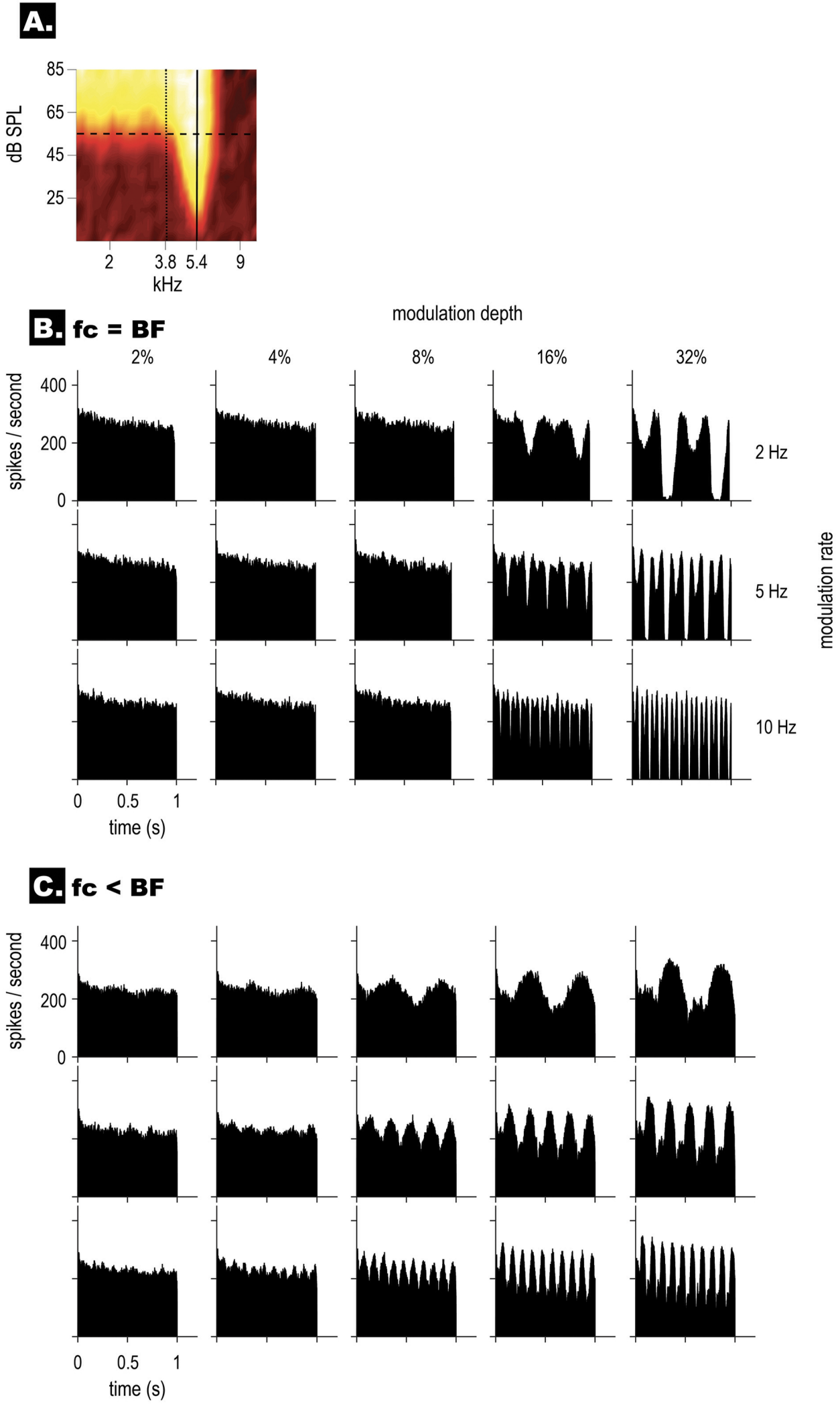

A.
spikes / bin
50
0
0
25
50
time (ms)
B.
dB SPL
85
55
25
4
8
15
kHz
C.
modulation depth
2%
4%
8%
16%
32%
spikes / second
500
250
0
2 Hz
5 Hz
10 Hz
modulation rate
0
0.5
1
time (s)
D.
Instantaneous frequency (Hz)
fc / BF
spikes/bin
0
0.5
1
cycle
Instantaneous frequency (Hz)
fc / BF
spikes/bin
sweep down
sweep up
E.
spikes / bin
80
20
*
0
0.5
1
cycle
F.
rate: 5 Hz
depth: 32%
SAC
XAC
normalized # concidences
3
2
1
0
-20
-10
0
10
20
delay (ms)
G.
Sumcor
depth: 32%
2 Hz
5 Hz
10 Hz
normalized # concidences
3
2
1
-10
0
10
delay (ms)
H.
Sumcor
rate: 5Hz
2 %
4 %
8 %
16 %
32 %
3
2
1
-10
0
10
delay (ms)

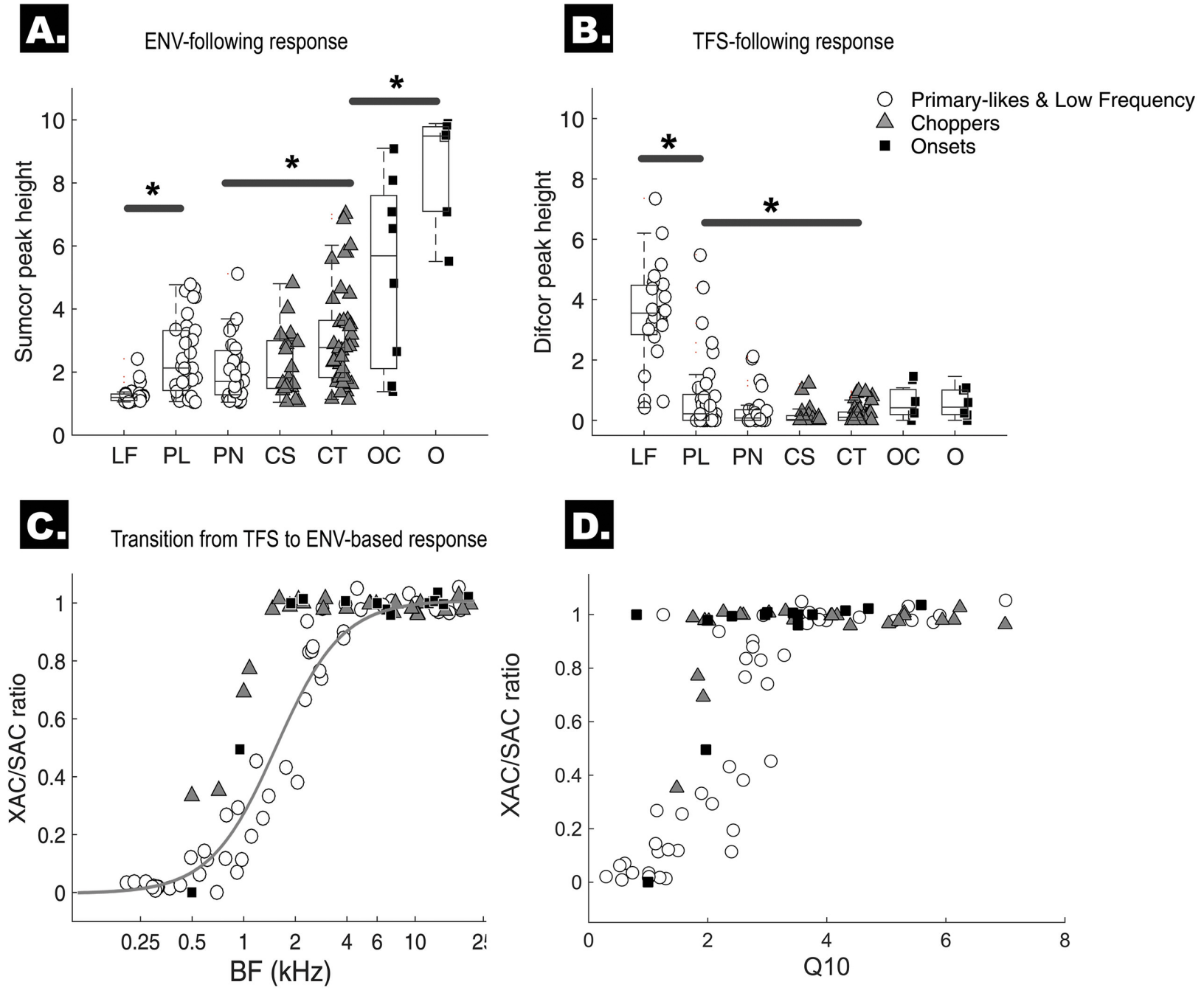

A.
spikes / bin
40
20
0
0
25
50
time (ms)
B.
dB SPL
85
55
25
0.1
0.3
1
kHz
C.
modulation depth
2 %
16 %
32 %
spikes / second
400
200
0
2 Hz
5 Hz
10 Hz
0
0.5
1
time (s)
D.
modulation rate
time (s)
1
0.5
0
0
10
20
interspike
interval (ms)
5
0
normalized
concidences
E.
rate: 5 Hz
depth: 32%
SAC
XAC
normalized
concidences
3
2
1
0
-20
-10
0
10
20
delay (ms)
F.
Difcor
depth: 32%
2 Hz
5 Hz
10 Hz
normalized
concidences
2
0
-2
-10
0
10
delay (ms)
G.
Difcor
rate: 5 Hz
2 %
4 %
8 %
16 %
32 %
2
0
-2
-10
0
10
delay (ms)

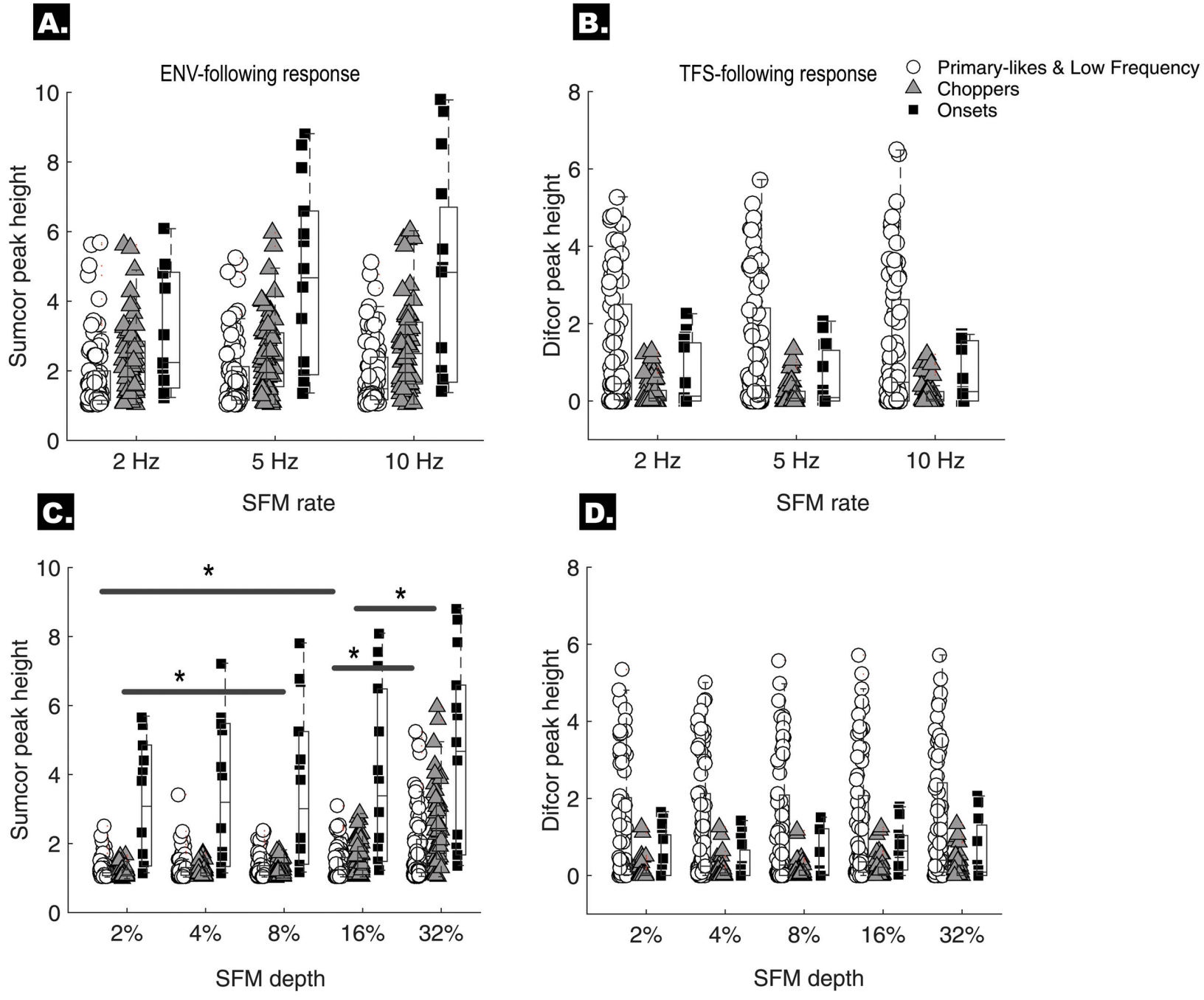

A.
spikes / bin
100
50
0
0
25
50
time (ms)
B.
dB SPL
85
55
25
3
7
15
kHz
C.
modulation depth
2%
4%
8%
16%
32%
spikes / second
400
200
0
75 dB
55 dB
35 dB
0
0.5
1
time (s)
level (dB SPL)
D.
Sumcor
rate: 5 Hz
depth : 32 %
35 dB
55 dB
75 dB
normalized # concidences
4
3
2
1
0
-10
-5
0
5
10
delay (ms)
E.
Sumcor - PL unit
rate: 5 Hz
depth: 32 %
35 dB
55 dB
65 dB
normalized # concidences
8
6
4
2
0
-10
-5
0
5
10
delay (ms)
F.
Difcor - LF unit
rate: 5 Hz
depth: 32 %
35 dB
55 dB
65 dB
4
2
0
-2
-4
-10
-5
0
5
10
delay (ms)

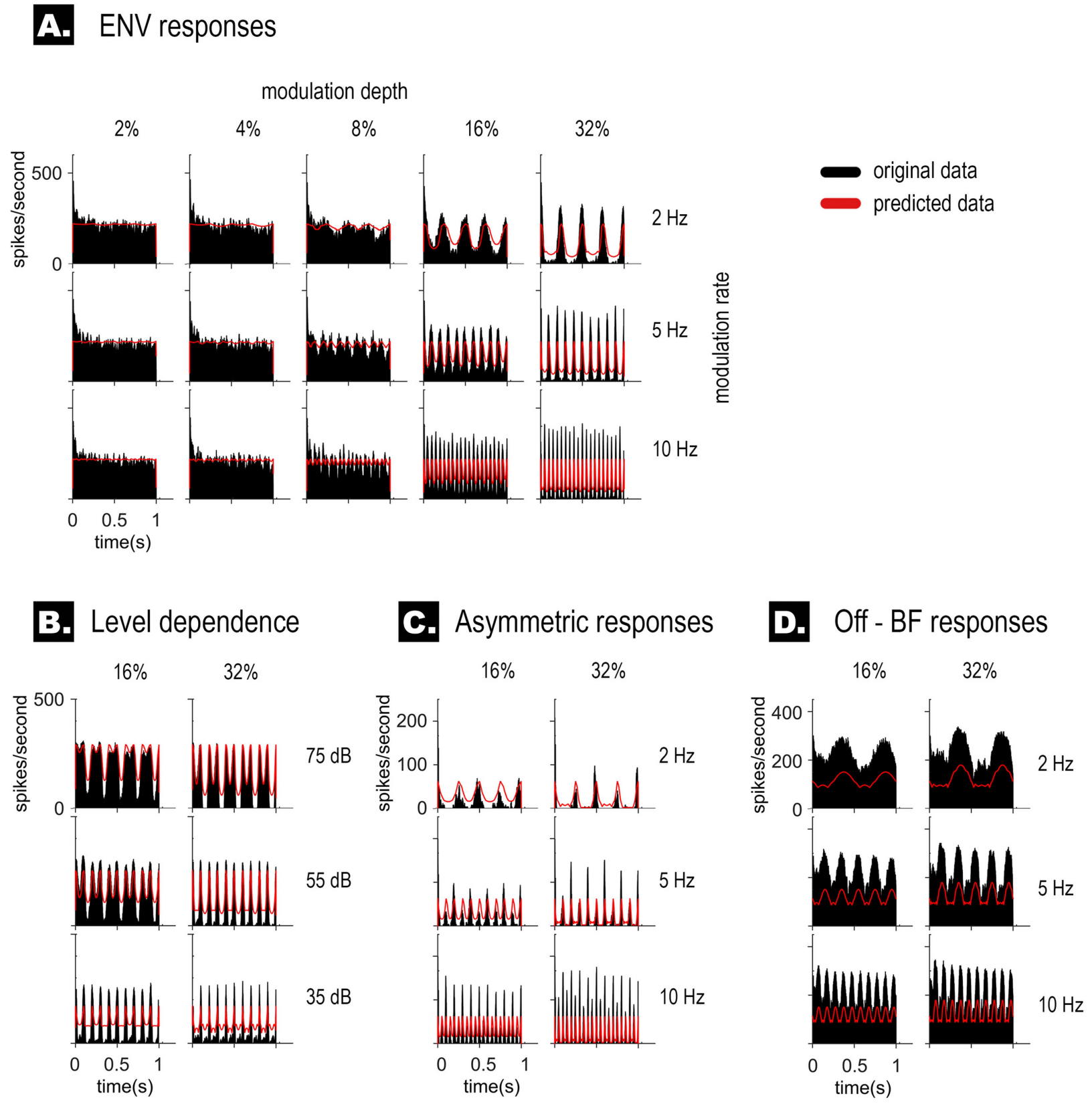

A.
200
100
0
spikes / bin
0
25
50
time (ms)
B.
85
55
25
dB SPL
3
6
12
kHz
C.
modulation depth
2%
4%
8%
16%
32%
200
100
0
spikes / second
2 Hz
5 Hz
10 Hz
modulation rate
0
0.5
1
time (s)
D.
spikes / bin
5
0
40
0
40
0
0
0.5
1
cycle
E.
PL, PN, LF
CS, CT
OC
OL
OI, O
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
DSI


A.
spikes / bin
100
0
0
50
time (ms)
B.
dB SPL
85
55
25
4
6
10
15
kHz
C.
spikes/second
200
100
0
0
0.5
1
time(s)


A.
85
65
45
25
dB SPL
2
3.8
5.4
9
kHz
B. fc = BF
modulation depth
2%
4%
8%
16%
32%
spikes / second
400
200
0
2 Hz
5 Hz
10 Hz
modulation rate
0
0.5
1
time (s)
C. fc < BF
spikes / second
400
200
0
0
0.5
1
time (s)


A.
ENV-following response
Sumcor peak height
LF
PL
PN
CS
CT
OC
O
B.
TFS-following response
Difcor peak height
Primary-likes & Low Frequency
Choppers
Onsets
C.
Transition from TFS to ENV-based response
XAC/SAC ratio
BF (kHz)
D.
XAC/SAC ratio
Q10


A.
ENV-following response
Sumcor peak height
10
8
6
4
2
0
2 Hz
5 Hz
10 Hz
SFM rate
B.
TFS-following response
Primary-likes & Low Frequency
Choppers
Onsets
Difcor peak height
8
6
4
2
0
2 Hz
5 Hz
10 Hz
SFM rate
C.
Sumcor peak height
10
8
6
4
2
0
*
*
*
*
2%
4%
8%
16%
32%
SFM depth
D.
Difcor peak height
8
6
4
2
0
2%
4%
8%
16%
32%
SFM depth


A.
ENV responses
modulation depth
2%
4%
8%
16%
32%
spikes/second
500
0
2 Hz
5 Hz
10 Hz
modulation rate
0
0.5
1
time(s)
original data
predicted data
B.
Level dependence
16%
32%
spikes/second
500
0
75 dB
55 dB
35 dB
0
0.5
1
time(s)
C.
Asymmetric responses
16%
32%
spikes/second
200
100
0
2 Hz
5 Hz
10 Hz
0
0.5
1
time(s)
D.
Off - BF responses
16%
32%
spikes/second
400
200
0
2 Hz
5 Hz
10 Hz
0
0.5
1
time(s)