BRAIN RESEARCH 1434 (2012) 90–101

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Modeling frequency modulated responses of midbrain auditory neurons based on trigger features and artificial neural networks

T.R. Chang[a], T.W. Chiu[b], X. Sun[c], Paul W.F. Poon[d,*]

[a]Dept. of Computer Sciences and Information Engineering, Southern Taiwan University, Tainan, Taiwan
[b]Dept. of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan
[c]Division of Life Sciences, East China Normal University, Shanghai, China
[d]Dept. of Physiology, Medical College, National Cheng Kung University, Tainan, Taiwan

ARTICLE INFO

Article history:
Accepted 21 September 2011
Available online 29 September 2011

Keywords:
Complex sound coding
Frequency modulation
Neural modeling
Spectro-temporal
Receptive field
Inferior colliculus

ABSTRACT

Frequency modulation (FM) is an important building block of communication signals for animals and human. Attempts to predict the response of central neurons to FM sounds have not been very successful, though achieving successful results could bring insights regarding the underlying neural mechanisms. Here we proposed a new method to predict responses of FM-sensitive neurons in the auditory midbrain. First we recorded single unit responses in anesthetized rats using a random FM tone to construct their spectro-temporal receptive fields (STRFs). Training of neurons in the artificial neural network to respond to a second random FM tone was based on the temporal information derived from the STRF. Specifically, the time window covered by the presumed trigger feature and its delay time to spike occurrence were used to train a finite impulse response neural network (FIRNN) to respond to this random FM. Finally we tested the model performance in predicting the response to another similar FM stimuli (third random FM tone). We found good performance in predicting the time of responses if not also the response magnitudes. Furthermore, the weighting function of the FIRNN showed temporal 'bumps' suggesting temporal integration of synaptic inputs from different frequency laminae.

This article is part of a Special Issue entitled: Neural Coding.

1. Introduction

Frequency modulation (FM) is an important building block of communication signals of animals and human (review see Kanwal and Rauschecker, 2007). Electrophysiological studies at the auditory midbrain and cortex of animals showed that many neurons are selectively sensitive to time-varying signals like FM tones (Atencio et al., 2007; Brown and Harrison, 2009; Eggermont, 1994; Heil et al., 1992; Poon et al., 1991; Qin et al., 2008; Rees and Møller, 1983; Whitfields and Evans, 1965; review see Suta et al., 2008). The exact neural mechanisms underlying FM coding remain somewhat elusive as their study often requires challenging techniques like *in vivo* whole cell patch clamp and the sample size is limited (Gittelman et al.,

* *Corresponding author*. Fax: +886 6236 2780.
E-mail addresses: trchang@mail.stut.edu.tw (T.R. Chang), twchiu@g2.nctu.edu.tw (T.W. Chiu), xdsun@bio.ecnu.edu.cn (X. Sun), ppoon@mail.ncku.edu.tw (P.W.F. Poon).

0006-8993/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2011.09.042

2009; Ye et al., 2010; Zhang et al., 2003). One approach to understand the neural mechanisms of FM coding is computational modeling of neural spike responses. This approach involves firstly constructing an input–output function for the cell to probe sounds, and then assessing its performance using test sounds. In modeling responses to complex sounds, results are only more satisfactory at the lower but not at the higher centers (Kim and Young, 1994; Lesica and Grothe, 2008; Reiss et al., 2007; Theunissen et al., 2000). This discrepancy in results is explained by a difference in their non-linear behaviors (Ahrens et al., 2008; Bar-Yosef et al., 2002; Christianson et al., 2008; Escabí and Schreiner, 2002; Young and Calhoun, 2005). It remains to be explored if the prediction of neural responses at centers higher than the cochlear nucleus (e.g., midbrain) can be improved through refinement of linear computational schemes, without incorporating substantial non-linearity in the model.

Along the ascending auditory pathways, the midbrain (or inferior colliculus) is an important center where the selective sensitivity to FM first emerges (Felsheim and Ostwald, 1996; Poon et al., 1991, 1992; Rees and Møller, 1983; review see Rees and Malmierca, 2006). Characteristically, these FM-sensitive neurons respond to tones of rapidly varying frequency, but they do not respond to pure tones. Computational attempts have been made to predict FM responses at the midbrain and cortex. Again, results are not satisfactory especially in predicting responses to naturally-occurring sounds. This poor prediction is not unexpected given the nonlinear behavior of central circuits (Aertsen and Johannesma, 1981b; Machens et al., 2004; Theunissen et al., 2000; Valentine and Eggermont, 2004; review see Escabí and Read, 2005).

One common modeling strategy makes use of the spectro-temporal receptive field (STRF; Aertsen and Johannesma, 1981a; Depireux et al., 2001; Eggermont et al., 1983; Hermes et al., 1981; Klein et al., 2000; Miller et al., 2002; review see Young, 2010). STRF is used to represent the input–output relationship of central neurons to sounds on the time-frequency plane. Typically, a probe tone of randomly varied frequency is used to evoke spike responses from an FM-sensitive cell (deCharms et al., 1998; Escabí and Schreiner, 2002; Poon and Yu, 2000; Theunissen et al., 2000). To construct the STRF, the analysis involves averaging the random sound energy preceding each spike. Band-like concentrations of energy appear in the STRF of FM-sensitive cells, at pre-spike intervals consistent with the neural transmission time measured from the auditory periphery to where the cell is recorded. These band-like structures form the putative 'trigger features' that determine the spike responses (analogous to the concept of receptive field in sensory physiology). In the STRF, trigger features of a variety of patterns have been reported (e.g., rising, falling orientations representing FM sensitivity, or a flat orientation representing the sensitivity to pure tones; examples see Atencio et al., 2007; Chiu and Poon, 2007). Often the exact pattern of trigger features depends on the property of stimuli used to generate the STRF (Valentine and Eggermont, 2004). The host of stimuli used to generate STRFs ranges from random tonal stimuli to naturally-occurring complex sounds (e.g., Escabí and Schreiner, 2002; Poon and Yu, 2000; Theunissen et al., 2000). No attempt has yet been made to incorporate temporal information from STRF into an artificial neural network in predicting FM responses.

In this study, we attempted to predict the cell's response to FM sounds using artificial neural network modeling based on two kinds of information derived from its STRF: (a) the 'time window' containing the trigger feature, and (b) the 'delay time' measured from the time of spike occurrence to the time of the trigger feature.

2. Results

Fig. 1 shows typical responses of an FM-sensitive cell to different sounds: a slow FM (tone-1, Fig. 1A, non-identical across trials), or a fast FM (tone-2 and tone-3, Fig. 1B, identical across trials). Note that the responses appear rather stationary across trials, and that the averaged time-locked response profiles (PSTHs) are minimally altered by the Gaussian filtering.

Most of the 10 FM-sensitive cells displayed strong response directionality, i.e., preferring one (instead of both) direction of modulation (Fig. 2). Within the pre-spike time window of about 5 to 29 ms, concentration of the modulating waveforms formed apparent 'hot bands' in the STRFs, typically situated above or below the carrier frequency, or also the best frequency (BF) of the cell. The orientation of the hot band represents a preference of individual neurons to the direction of frequency modulation (8 cells to up-sweeps, and 2 cells to down-sweeps). The average time interval between the proximal end of the hot band and the time of spike occurrence is 8.04 (±3.93 SD) ms, which represents the central transmission delay from the cochlea to the spike generating mechanism of the cell from which the spikes were recorded. These FM-sensitive cells (or sometimes called 'FM-specialized' cells) represented about 1/3 of the click-responsive neurons in the rat auditory midbrain (Chiu and Poon, 2007; Poon and Chiu, 2000; Poon et al., 1991, 1992). These 10 neurons were selected (from a population of 15 FM-sensitive cells) based on their STRFs which displayed a simple trigger feature. For simplicity, we did not include neurons with complex STRFs (e.g., containing both up and down FM sweeps, and/or with multiple FM sweeps, an example is shown in Supplementary Fig. 1). Because of the temporal differences between the slow FM (tone-1) and fast FM (tone-2, tone-3), for a given cell, the exact trigger features in the respective STRFs appeared non-identical, though they tended to fall within similar if not overlapping time windows (Fig. 2 left column versus right). Slow FMs, in comparison with fast FMs, usually produced more clearly outlined trigger features with slightly longer durations. This is likely more related to the difference in modulating waveforms than in spike counts (as spike counts are not statistically different between the FM datasets, Wilcoxon signed rank test, $P=0.14$).

For these 10 cells, our model showed highly satisfactory performance in learning spike responses to the training dataset. The match between the trained and actual PSTHs is typically high when expressed in terms of average overlap between the two PSTHs (96.68±1.64%, mean±SD, n=10). After training, the prediction of the response to tone-3 was almost as high (96.31±1.75%, mean±SD, Fig. 3). Swapping the training and testing datasets made little difference in the prediction performance. Results indicated that the current method of using 1 second long training data and the training iteration

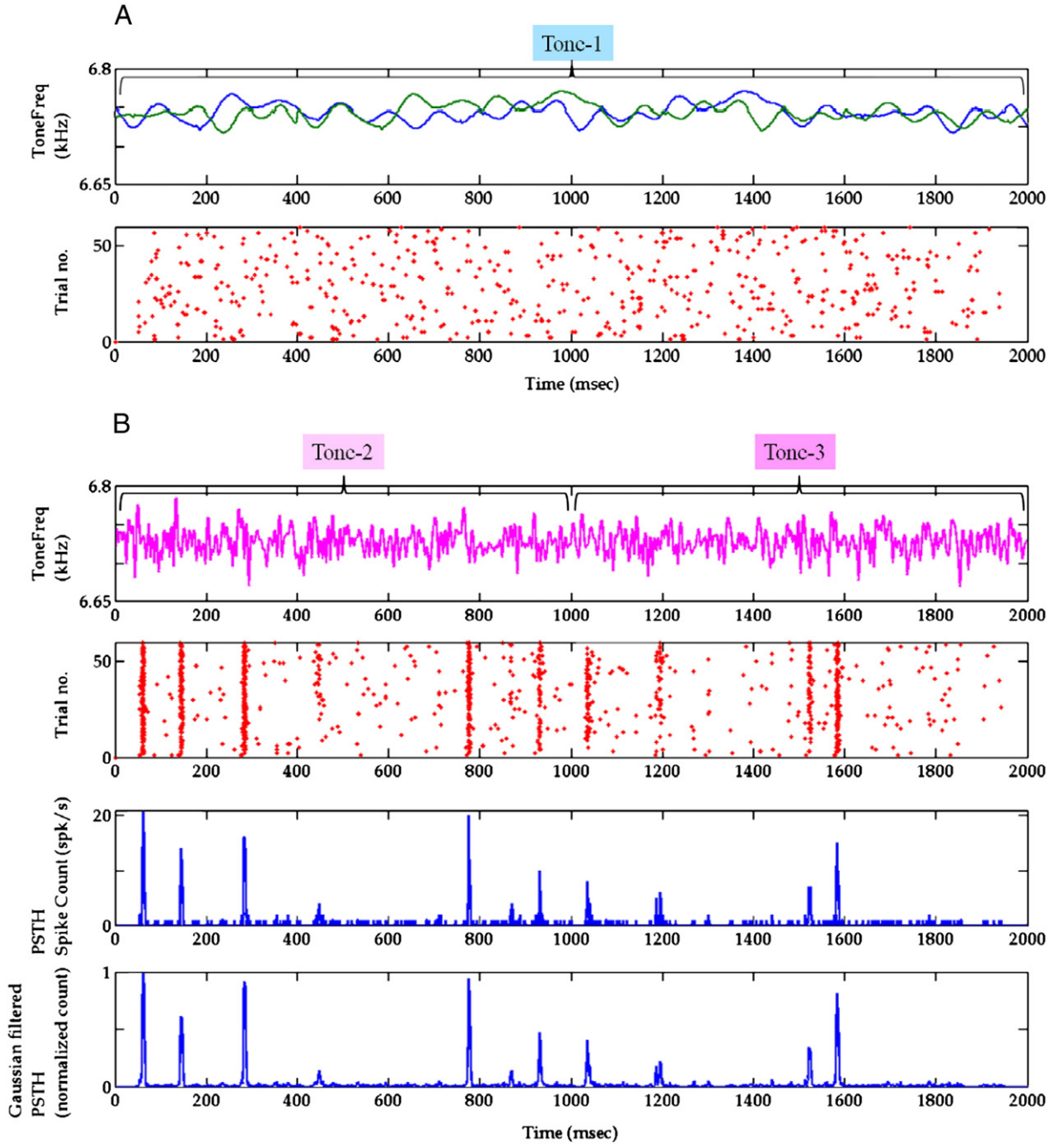


Fig. 1 – Responses of an FM-sensitive cell: (A) the slow FM stimulus (tone-1) shown by two superimposed instantaneous frequency time-profiles (top panel), and the corresponding spike responses in dot-raster (bottom, each dot represents an action potential); (B): similar plots for FM tone-2 (top panel, left) and tone-3 (top panel, right) with the original PSTH and the Gaussian filtered version (bottom 2 panels). The time-locked responses reflect activation by the presumed trigger features in the stimulus. Note that half of the dataset was used for training of the model, and the remaining half for testing.

of 500 predicts satisfactory results with minimal effect of over-fitting.

While the model was tested on datasets with spike jitters, trigger features in the STRF were determined after a procedure of de-jittering. Therefore, the temporal information derived from the trigger feature could be slightly off-optimal. To determine how critical the time window and delay time are in the model performance, a systematic scanning of both the time window length and delay was carried out. This procedure scans results systematically across 'delay time' and 'time window' (delay time range: 1 to 16 ms; time window range: 1 to 20 ms). The performance of the FIRNN model at each combination of 'delay time' and 'time window' is then represented by a color-coded pixel on a 2D plot (with y axis representing 'delay time', and x-axis 'time window'). Here, because the performance assessed by index-1 was typically high, a stricter index (index-2) was used for comparison. Index-2 takes into account only the part of PSTH showing spike responses

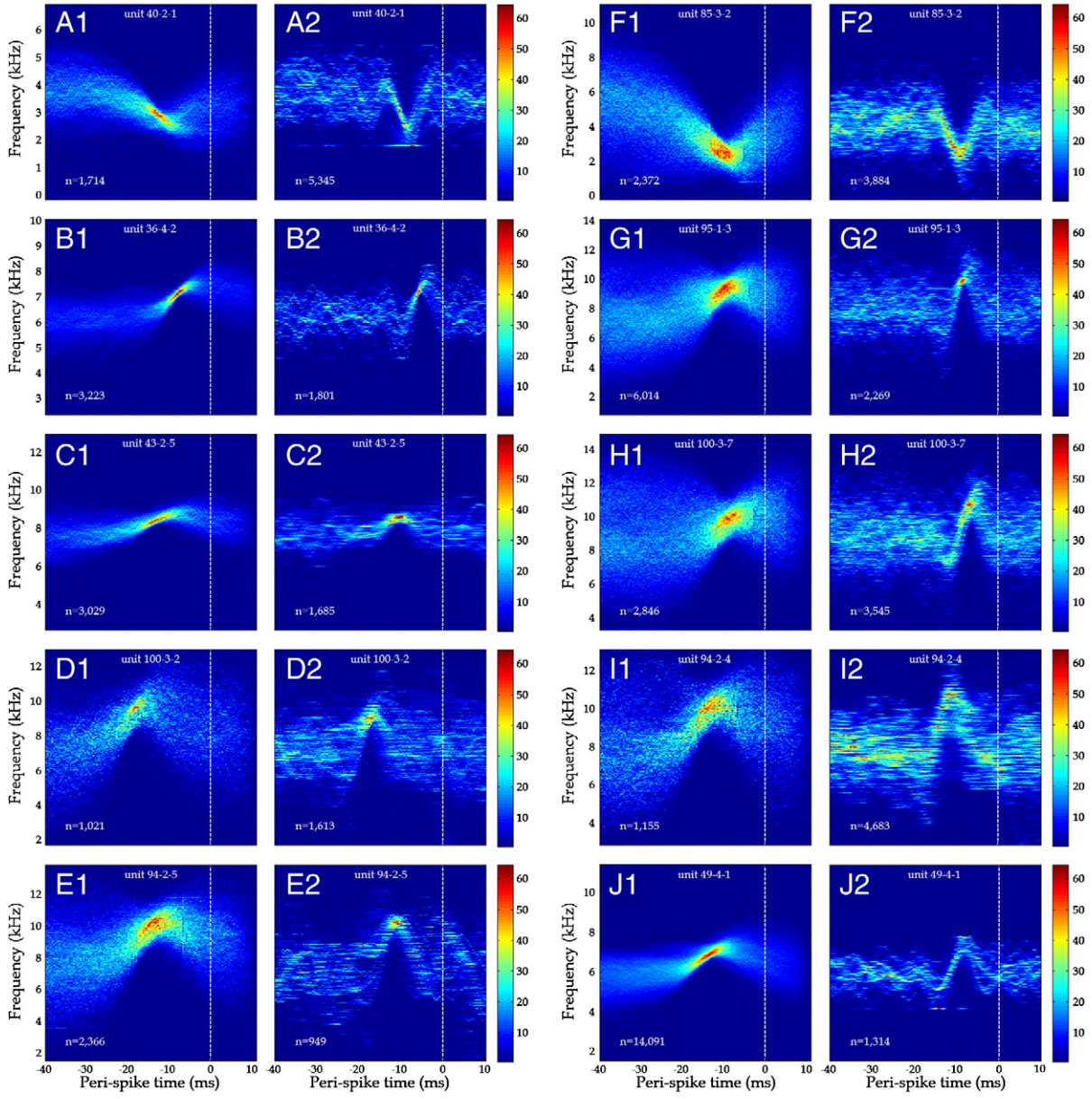


Fig. 2 – STRFs of 10 FM-sensitive cells in response to slow FM (tone-1; left columns: A1 to J1) and fast FM (tone-2 and tone-3; right columns: A2 to J2). In this and other similar plots, the overlap of frequency time-profiles at each pixel is color coded, with 'hot' areas representing high overlaps or the trigger features. In each panel, 'n' is the number of stimulus time profiles, and the spike occurrence marked by the vertical dashed line.

(actual and predicted) without considering the part showing no response (see Experimental procedures). We noted only a weak correlation between the two performance indices (P=0.30, Pearson correlation, n=10), likely because of the appearance of different silent periods in the PSTH across cells. Fig. 4 shows the results from all 10 cells (white arrows mark the time window and time delay estimated from STRF trigger features). In most cases (8 out of 10 cells), arrows are found inside the region of high performance (>90% of peak: red areas enclosed by white contours). In each panel, the prediction level drops off at long delays (top, blue areas), suggesting the existance of a critical delay time. At long time windows, performance reached a plateau-like level, implying also the existance of a critical time window. Results showed that when modeling at a fixed delay time, better performance was always associated with longer time windows, and such results were not expected if cells responded only to a steady tone. At excessively short time windows and short delay times, the model performance became invariably poor. Within each red area enclosed by the contour, the model performance was typically high and often showed striking similarity with the PSTH plots (Supplementary Fig. 2).

For simple artificial neural networks (with single neuron in the hidden layer) like the present one, it is possible to examine in greater details the weighting functions obtained at different combinations of time window and delay time. Fig. 5 (and Supplementary Fig. 3) shows the various weighting functions within the region of high performance (Fig. 4). Some

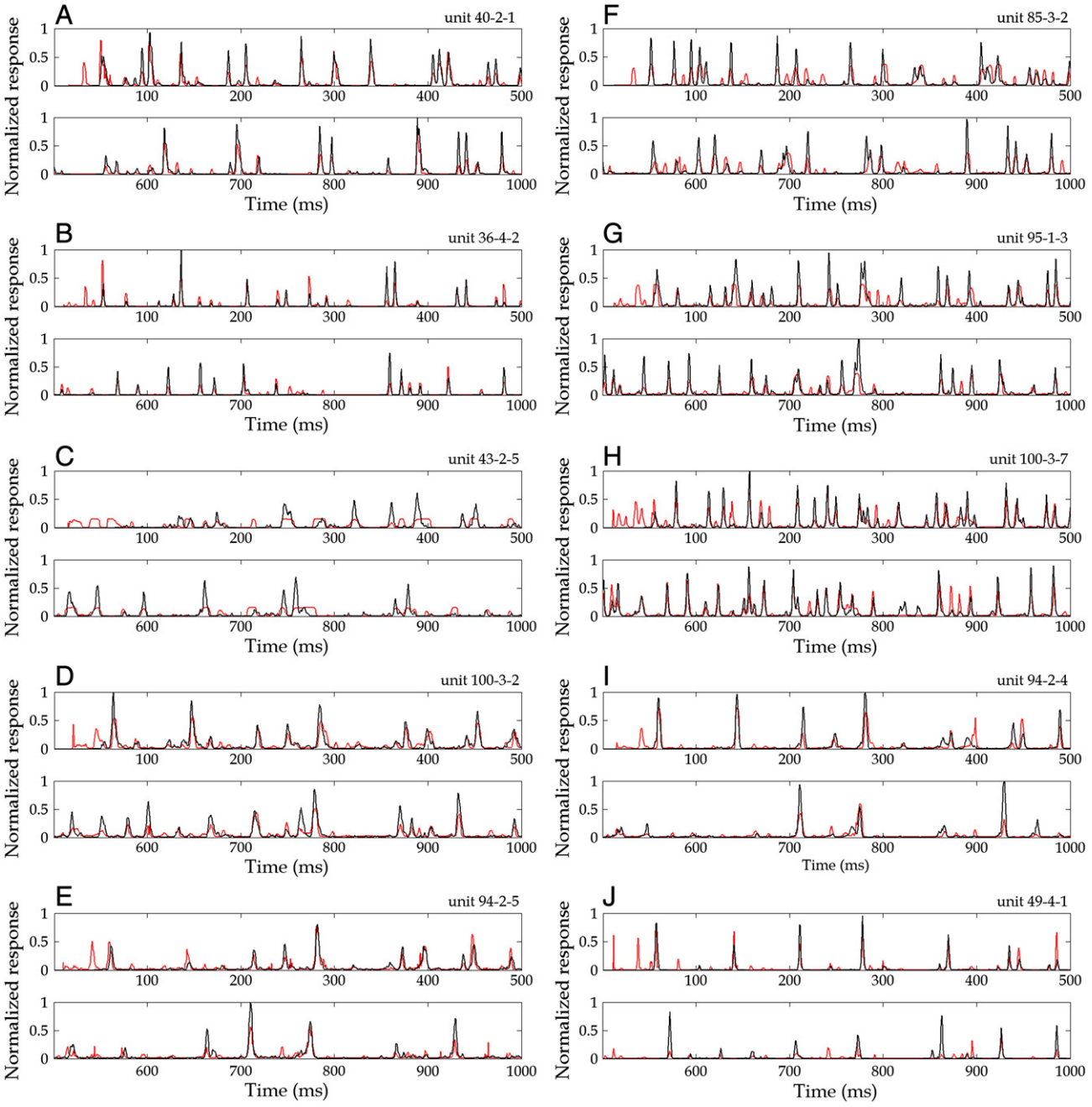
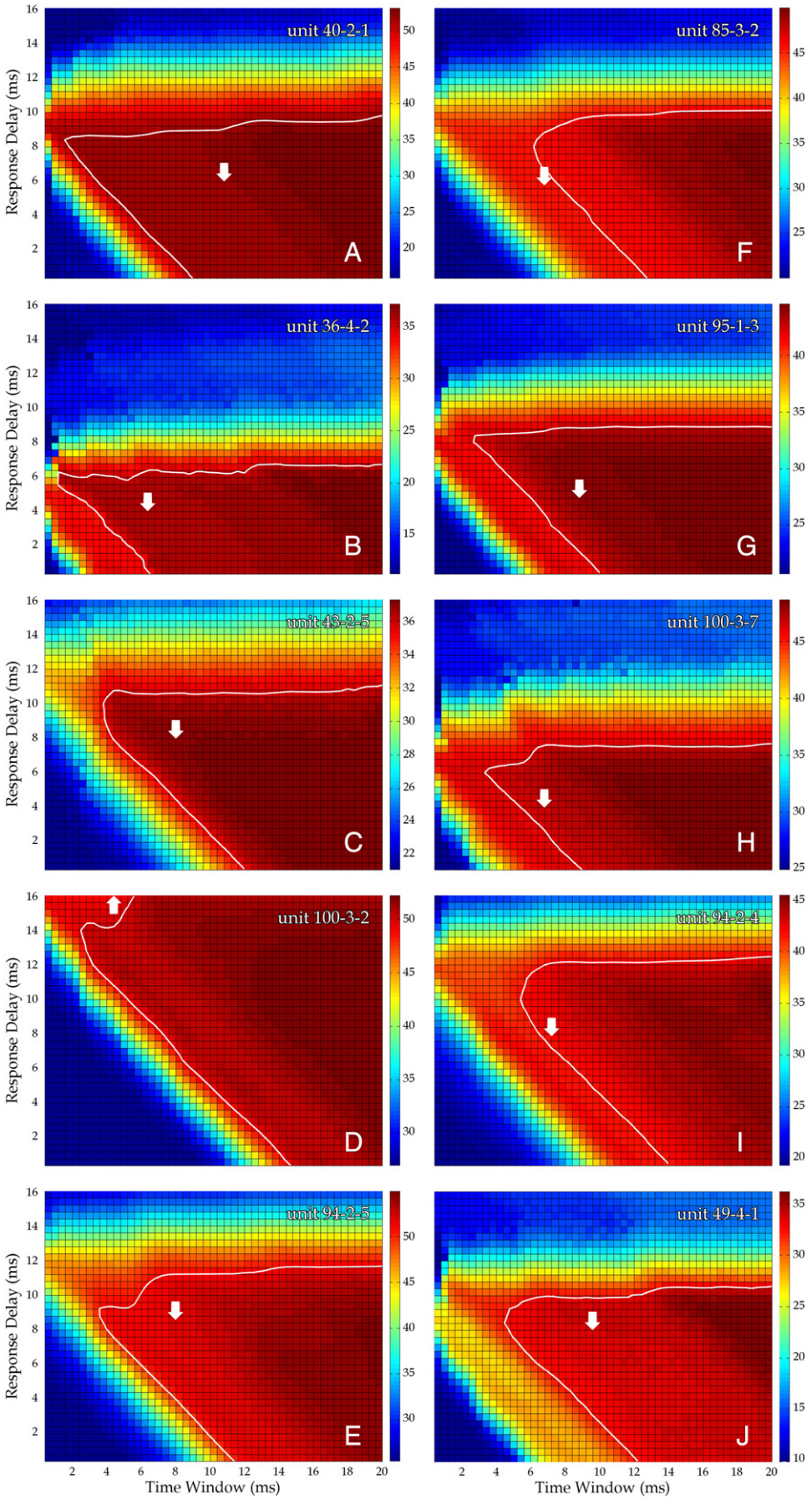


Fig. 3 – Model prediction of PSTH responses in 6 FM-sensitive cells showing the actual (black) and predicted results (red). To visualize details, responses to the 1 s long stimulation are displayed in 2 time segments. Note the large overlaps. Cell labels correspond to those in Fig. 2.

consistent features are apparent. Firstly, heavily weighted elements (or 'bump-like' structures) are consistently observed in the weighting function, particularly for cells with a preference to the FM up-sweeps. In their STRFs, the 'bumps' appeared like 'valleys' in cases of preference to a down-sweep. These features are more clearly seen at long time windows and short delay times. In general, the duration of the 'bumps' (or 'valleys) tended to overlap with the time window of trigger feature, suggesting the importance of trigger features in the modeling. Secondly, at certain combination of

Fig. 4 – Model performance (percentage of overlap in PSTH responsive area) for the 10 cells plotted as the results of systemically scanning the length of time window and the delay time. Arrows indicate the time window and delay determined based on trigger features in the STRF. Note the areas of peak performance determined by systematic scan (dashed regions) are close to if not inclusive of the arrows.


unit 40-2-1
A
unit 85-3-2
F
unit 36-4-2
B
unit 95-1-3
G
unit 43-2-5
C
unit 100-3-7
H
unit 100-3-2
D
unit 94-2-4
I
unit 94-2-5
E
unit 49-4-1
J
Response Delay (ms)
Time Window (ms)

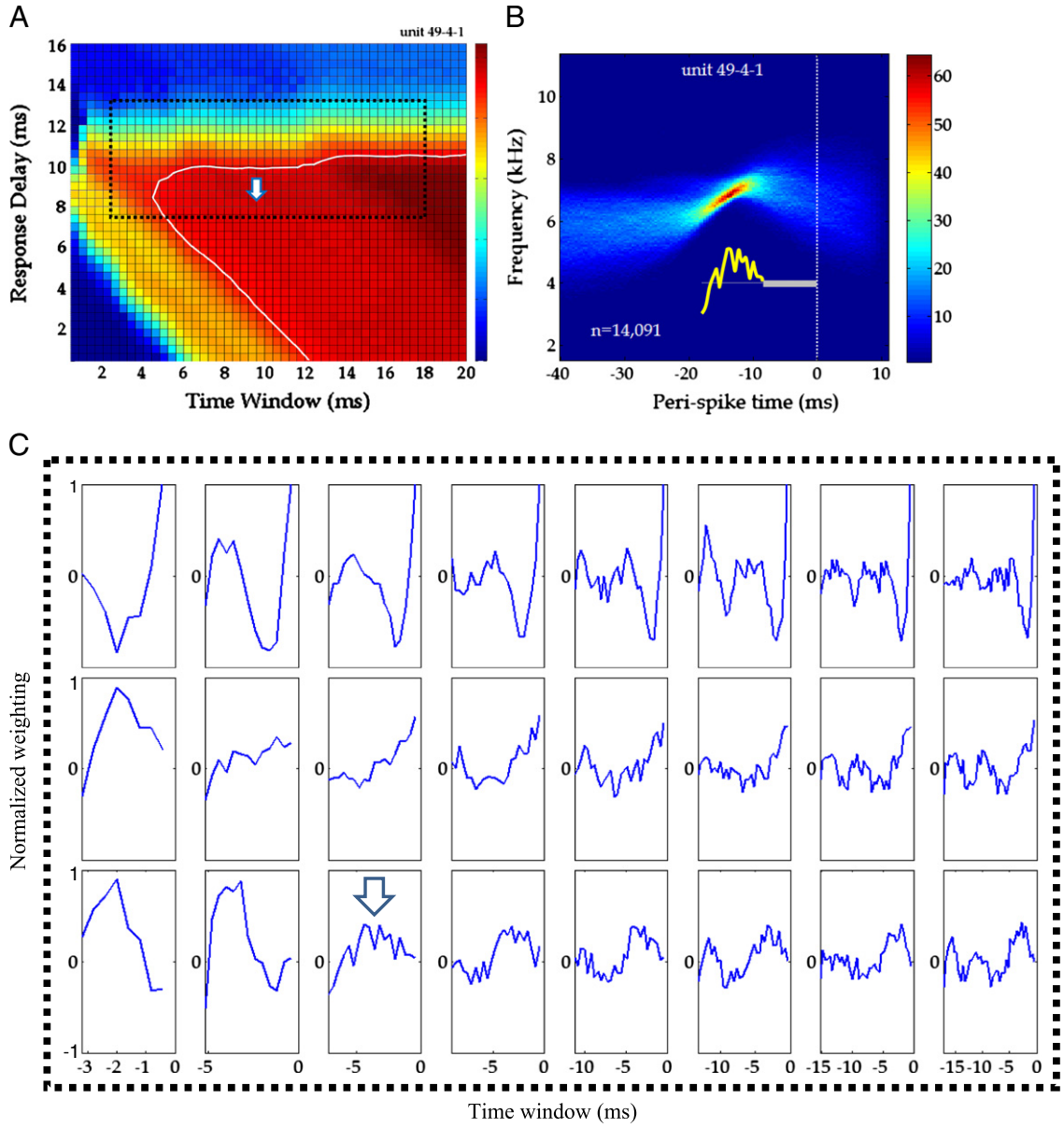


Fig. 5 – Weighting functions trained at different combinations of time window and delay time on a cell. (A) Same plot of panel J in Fig. 4, showing the region of systematic scan (in black dotted rectangle) where the various weighting functions are displayed in (B). Note the systematic scan produces more weighting functions with finer steps. But for simplicity, only part the results (taken at coarser steps) is shown in (B). Note also the time axis is minus as it refers to the time series in the FIRNN, with zero closest to the spike occurrence. Arrows indicate the time window and delay time based on STRF trigger features. (C) Relationship of the weighting function (indicated by the arrow in B) to the STRF. Yellow curves: weighting functions; thin gray line: zero; thick gray line: delay time. Note weighting function overlaps with the trigger feature, and its multiple-peak appearance. Examples of weighting functions from other cells are shown in Supplementary Fig. 3.

time window and time delay near that of the trigger feature, multiple peak-like structures were often observed riding on the 'bumps' (or 'valleys'), suggesting synaptic activations.

3. Discussion

Results showed that the stimulus energy within a critical pre-spike time window is important for successful prediction of the output of FM-sensitive neurons displaying simple trigger features in their STRFs. One implication is that for these FM-sensitive cells, the observed response is mainly generated by excitation, provided that the stimulus energy has passed though the spectro-temporal domain of the trigger feature. The finding of primarily excitation rather than inhibition in our study could be related to the selection of cells with simple STRFs for modeling, and the use of a narrow-band FM stimulus presented at a relatively low sound intensity (~30 dB supra-threshold). Neurons which show only excitations in STRF in response to narrow band sounds of low intensities could show

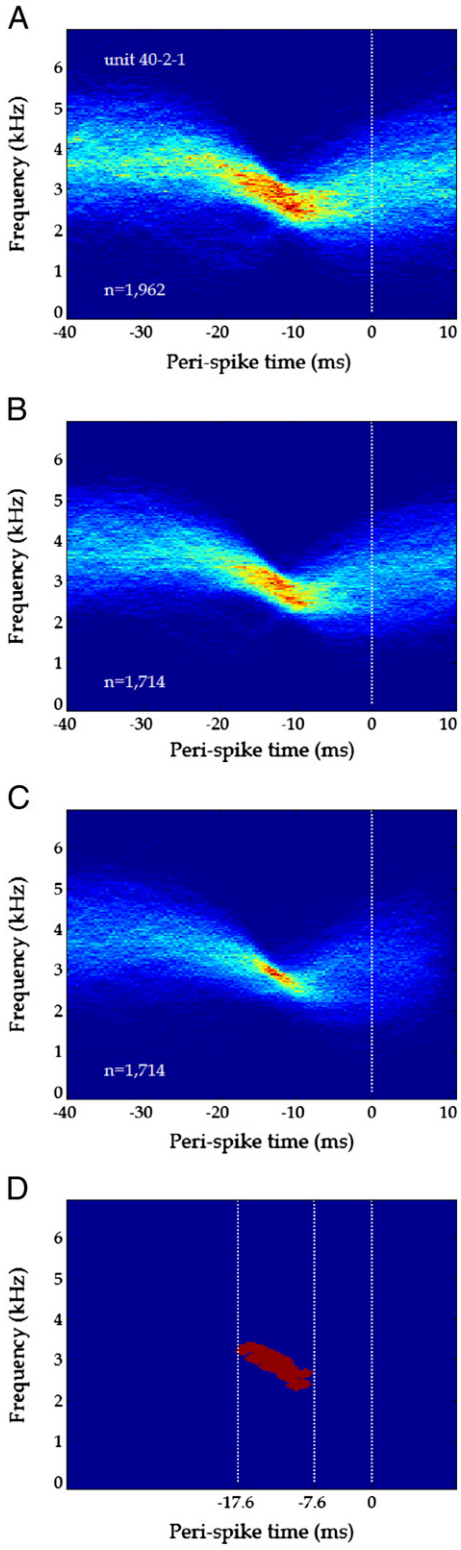


Fig. 6 – Example of an FM-sensitive cell illustrating how temporal information is derived from the STRF. (A) STRF constructed through spike-triggered averaging; (B) main trigger feature extracted with the progressive thresholding method (details see Supplementary Fig. 4); (C) STRF after spike de-jittering; (D) the outline of extracted trigger feature at the critical threshold showing the start and end times (–17.6 and –7.6 ms). The central transmission delay time is 7.6 ms in this case.

additional inhibitions when stimulated with wide-band sounds at higher levels (review see Eggermont, 2011). The finding that a pre-spike time window contains trigger features is well consistent with the literature on FM cells (Chiu and Poon, 2007; Qiu et al., 2003). The artificial neural modeling approach, due to its self-adjusted weighting function (especially in the case of systematic scanning), avoids the constraint of a fixed trigger feature which is present in the case of modeling with STRF convolution (Christianson et al., 2008; Kao et al., 1997). Evidently, as long as the trigger features are not too complex, our model predicts the response rather satisfactorily even with training datasets as brief as 1 second long, without involving substantial non-linear elements in the network.

The temporal information in the STRF likely reflects two underlying mechanisms: (a) a more or less fixed central transmission delay time (i.e., for neural transmission to travel from the inner ear to where spike activity is recorded), and (b) a more or less fixed synaptic processing time window (i.e., for the different synaptic events to take place over its dendrites following the arrival of neural inputs). Biologically relevant information may be derived from the weighting function of the FIRNN model in some cases. Temporal segments of the pre-spike signal are heavily weighted (appearing as 'bumps' or 'valleys' in the weighting function). The multiple-peak appearance at the segment of heavy weightings could resemble synaptic excitations occurring in a temporal sequence, likely corresponded to synaptic inputs from nearby 'iso-frequency laminae'. This temporal integration of synaptic inputs across frequency laminae could be a mechanism for the detection of modulation direction. Our current result is different from the finding on FM directional sensitivity at the midbrain of the echolocating bat, where amplitude asymmetry of postsynaptic potentials appears more important than temporal integration (Gittelman and Pollak, 2011; Gittelman et al., 2009). The difference could be related to limitations of the *in vivo* patch clamp technique which tends to reveal synaptic events closer to the soma than those at the dendrites. Also, one cannot exclude the possibility of different mechanisms of FM directional sensitivity that exist across species especially for auditory-specialized animals like the bat. Future experiments using intracellular recordings from rat FM-sensitive neurons will help solving the controversy.

While the positive weight values are likely more consistent with excitation, the negative values should be viewed with more caution, as they cannot be considered simply as inhibition. Since during training, the FIRNN estimated the PSTH profile which contained periods of both spike responses and no responses. The weighting function of the trained FIRNN was then convolved with the modulating time waveform of the test dataset (expressed as positive and negative values with respect to the carrier frequency) to predict the response. What the model had to consider was apparently: (a) the temporal features of the frequency sweep (up- or down-FM, its rate of change, and the extent of modulation), and (b) which side with respect to the BF of the cell the FM stimulus had occurred (i.e., frequency range above or below the BF). In other words, one may consider that the FIRNN was basically modeling the integration of synaptic inputs that landed on dendritic field of the cell traversing 'iso-frequency laminae' likely on either side of the cell body (Poon et al., 1992).

Inhibition is known to be involved in FM detection (Fuzessery and Hall, 1996; Gittelman et al., 2009) and the auditory midbrain is known to be rich in inhibitory mechanisms (including post-

inhibitory rebound; Sivaramakrishnan and Oliver, 2001; review see Casseday et al., 2002). It is conceivable that some inhibitory influences might have been taken care of by the self-adjusted FIRNN weightings. But the involvement of inhibition especially at higher stimulus levels remains to be explored. As the spike occurrence determines when to capture the pre-spike modulating

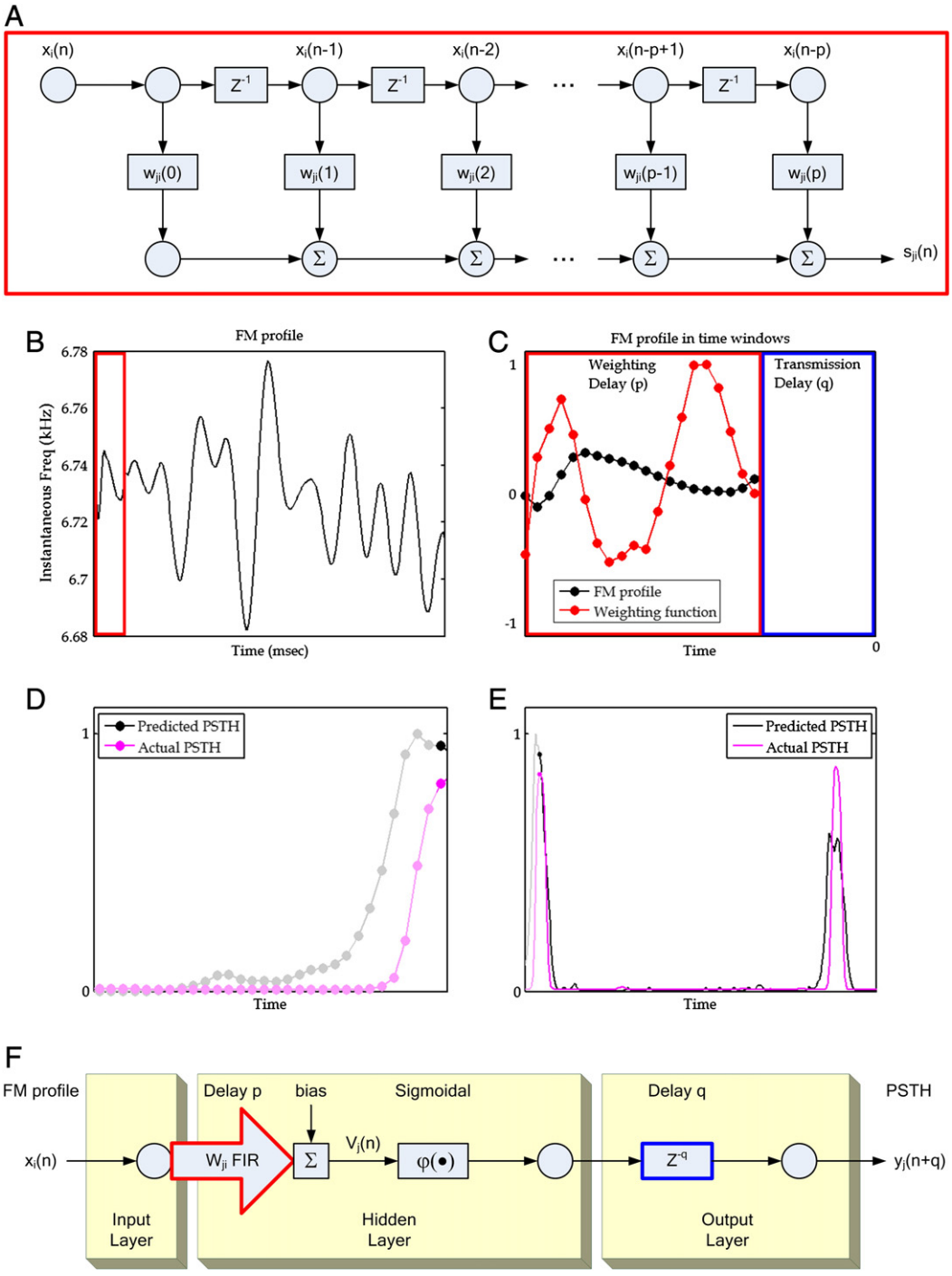


Fig. 7 – (A) Schematic diagram showing the structure of 'finite impulse response neural network' (FIRNN) of a hidden neuron. (B) The instantaneous frequency time profile of the stimulus (or modulating time waveform; black tracing) is fed to the input neuron, each time with a segment of input waveform (red rectangled area). The length of time segment is based on the STRF trigger feature (in response to tone-1). (C) After training (with the response to tone-2), the weighting function (red symbols) of the hidden neuron is convolved with the test stimulus (tone-3, black symbols), and the output is generated after a fixed delay (Delay (q), a quantity also determined by the STRF trigger feature to tone-1). (D) The temporal outputs of the model (gray or black symbols) are shown together with the actual responses (pink symbols). (E) A similar plot as in D but showing extended results over time. (F) Overall architecture of the FIRNN that includes three neurons: the input, hidden and output. For details see supplementary text.

waveform, periods of 'inhibition', which appears indistinguishable from periods of no response, cannot be studied with this spike-averaging approach.

Major differences exist between a previous FIRNN model (Chang et al., 2003) and the present one. First, the current method has a simple architecture (3 layers, with 1+1+1, or a total of 3 neurons from the input to output layers). In contrast, the previous model has a more complicated architecture (4 layers, with an architecture of 1+10+10+1, or a total of 22 neurons from input to output with somewhat arbitrary delays times), and it also needs to divide the PSTH peaks into strong and weak responses (again at somewhat arbitrary peak levels) for separate modeling. More importantly, the previous approach is a phenomenological model. Secondly, the present model is based on biologically relevant information in the STRF (assuming spectral integration of inputs over a specific time window and the output is generated after a transmission delay). In contrast, the previous model yields a total of 20 weighting functions for neurons in the 2 hidden layers, and the functions are harder to relate to the underlying biology. Thirdly, the present model finishes computation typically within 30 min, compared to about 2 days with the previous model. Perhaps due to its simplicity, the performance of the current model was less satisfactory when applied to predict responses of FM-sensitive cells with complex STRF (e.g., Supplementary Fig. 1). For some cells, when more neurons were added to the first layer of the FIRNN, prediction markedly improved in performance, including responses that resembled post-inhibitory rebounds (see Supplementary Fig. 4). This advantage with additional neurons in the hidden layer of the FIRNN opens up the possibility in differentiating mechanisms that generate different categories of response in the PSTH.

While the present model has performed satisfactorily for 'in-group' stimuli (i.e., tone-2 and tone-3), it remains to be determined how well it predicts the responses to other complex sounds (Eggermont et al., 1983; Gourevitch et al., 2009; Machens et al., 2004; Rodríguez et al., 2010; Yeshurun et al., 1989). It is now known that fine or complex trigger features are present in the STRF (Atencio et al., 2009; Chang et al., 2010a, 2010b; Chiu and Poon, 2007) as revealed after spike de-jittering (Chang et al., 2005; Gollisch, 2006; Linden et al., 2003; Versnel et al., 2009). Modeling responses to individual features of FM in the complex sound with separate FIRNNs, and adding neurons to the hidden layer in a systematic way would be useful in predicting the responses of more auditory neurons.

4. Experimental procedures

Extracellular single unit responses of auditory midbrain neurons to sound stimuli were first recorded in urethane-anesthetized rats (Sprague Dawley strain, 250–350 g b.w.) using 3 KCl-filled glass micropipette electrodes (30–80 MΩ) according to conventional electrophysiological procedures as we had reported earlier (Chang et al., 2010a; Chiu and Poon, 2007). Units were mostly recorded from the central and external nuclei of the inferior colliculus, as judged by the recording depth, response latency and the tonotopic progression during electrode passage down the midbrain. Sounds were delivered from a free-field speaker placed 30° off-midline, 70 cm away, contralateral to the side of recording. Two sound stimuli of different statistical properties were presented (at ~30 dB suprathreshold: (a) random FM (designated tone-1) that is different across 60 repeated trials (2 s/trial, 0.2 s inter-trial interval) (Fig. 1A), and (b) random FM (tone-2 and tone-3) that is identical across trials (Fig. 1B). The bandwidth of the FM tone was typically <2.5 octave with respect to the BF of the neuron (this covered the spectral range of >90% of FM cells in the auditory midbrain; see Poon et al., 1991). The tone frequency was varied around the BF in a basically random fashion. The stimuli were obtained first by low-pass-filtering a white noise signal and the result was used to modulate the instantaneous frequency of a time-varying tone (through the voltage-control-frequency function of a sine wave generator) to produce the random FM signal. The two FM stimuli differ in the low-pass cut-off-frequency (25 Hz for tone-1, 125 Hz for tone-2 and tone-3). For better model performance, the slow FM was used to generate STRFs with trigger features spanning a longer time period. For each recorded cell, FM tone-1 was used to generate the STRF, FM tone-2 for model training, and FM tone-3 (identical in statistical property with tone-2, but differs in exact modulating time waveform; they are also known as 'in-group'stimuli) for model testing. PSTHs were generated by counting spikes/60 trials at 0.4 ms time bins, followed by low-pass filtering (multi-scale Gaussian filter: mean=0, variance=5; details see supplementary text) to represent the response strength.

STRFs were generated by adding, on the spectro-temporal plane, the instantaneous frequency time profiles within 40 ms before and 10 ms after the spike (Fig. 6A), followed by a procedure of de-jittering (Chang et al., 2005) to enhance the trigger features (Fig. 6B). The STRF was analyzed with a procedure of 'progressive thresholding' (Chang et al., 2010b; details see Supplementary Fig. 5) to extract the 'main' trigger feature (Fig. 6C). The main trigger feature of FM-sensitive cells typically appeared as a single band-like structure in the STRF. The time window covering the main trigger feature was then determined (with the 'time window' length equals to the difference between the start and end times of the trigger feature, and the end time with respect to the spike occurrence gives the 'delay time'; Fig. 6D). The time interval between the spike occurrence and the proximal end of the trigger feature is known to be close to the 'central transmission delay' (Chang et al., 2010a; Poon and Yu, 2000). Using the temporal information from STRF trigger features, the artificial neural network was trained to use the instantaneous frequency time-profile of a training dataset to estimate the empirical responses in the PSTH (Fig. 7). A multi-scale Gaussian function (mean=0, variance=5) was again convolved with the model output to simulate synaptic activation of the cell and to improve predictive power (Englitz et al., 2010). In greater details, a 'finite impulse response neural network' (FIRNN; Back et al., 1994) which models synaptic interactions as FIR linear filters was constructed in the form of an autoregressive time series (Fig. 7A,F; see Supplementary text). After training (500 iteration with tone-2 dataset), the FIRNN model acquired an optimized weighting function at a given delay time (Fig. 7C). The performance of the model to predict PSTH responses of a testing dataset (tone-3) was finally assessed with a single iteration. As shown in Fig. 7, the modulating time waveform (Fig. 7B) which represents the tone frequency of the FM stimulus is fed to the

neuron in the input layer. The receiving neuron in the hidden layer convolves its trained weighting function with the input instantaneous frequency profile (Fig. 7C). The frequency profile signal is normalized between + and − 1.0, to represent the frequency range above and below the BF of the cell. The convolution process produces an output, which is further biased (to cater for silent periods in the PSTH), and transformed with a sigmoidal function (optimized to simulate the non-linear response-level function of the cell). Before the final output, the predicted value is further delayed at the output layer (to cater for the central transmission time). Hence, for each step of processing based on the sliding window, a single point on the output PSTH is generated to represent the instantaneous strength of spike discharges (Figs. 7D, E).

Two indices were used to assess performance of the model: (a) index-1 'percentage of overlap in PSTH': we first compute the average percentage-disparity between the actual and predicted PSTHs (mean absolute difference in y-values of all time points), and then subtract the result from 100%, or (b) index-2 'percentage of overlap in the PSTH responsive area': we compute the area of positive hit ('intercept') and divide it by the total area ('union') of actual and predicted responses. Index-1 is what the FIRNN was trained to fit (to respond to some stimulus features and not to others), and index-2 is a stricter criterion as it considers only the areas of PSTH with spike responses and ignores those without. Since the FIRNN was trained to estimate the PSTH based on both positive and negative hits, index-1 always gives a better performance. The alternative index based on correlating time points between the predicted and actual PSTHs would yield performance levels intermediate between those generated by the present 2 indices.

Experimental procedure was approved by the Animal Ethics Committee of NCKU Medical College, Taiwan.

Acknowledgments

We thank Prof. Iain C. Bruce for editing the manuscript, and the three reviewers for their constructive comments. This research was supported by grants from National Science Council, NSC-99-2221-E-218-007, NSC-100-2923-B-006-001,Taiwan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.brainres.2011.09.042.

REFERENCES

Aertsen, A.M., Johannesma, P.I., 1981a. The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol. Cybern. 42, 133–143.

Aertsen, A.M., Johannesma, P.I., 1981b. A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. Biol. Cybern. 42, 145–156.

Ahrens, M.B., Linden, J.F., Sahani, M., 2008. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. J. Neurosci. 28, 1929–1942.

Atencio, C.A., Blake, D.T., Strata, F., Cheung, S.W., Merzenich, M.M., Schreiner, C.E., 2007. Frequency-modulation encoding in the primary auditory cortex of the awake owl monkey. J. Neurophysiol. 98, 2182–2195.

Atencio, C.A., Sharpee, T.O., Schreiner, C.E., 2009. Hierarchical computation in the canonical auditory cortical circuit. Proc. Natl. Acad. Sci. 106, 21894–21899.

Back, A., Wan, E., Lawrence, S., Tsoi, A.C., 1994. A unifying view of some training algorithms for multi-layer perceptrons with FIR filter synapses. IEEE Workshop on Neural Networks and Signal Processing, pp. 146–154.

Bar-Yosef, O., Rotman, Y., Nelken, I., 2002. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J. Neurosci. 22, 8619–8632.

Brown, T.A., Harrison, R.V., 2009. Responses of neurons in chinchilla auditory cortex to frequency-modulated tones. J. Neurophysiol. 101, 2017–2029.

Casseday, J.H., Fremouw, T., Covey, E., 2002. The inferior colliculus: a hub for the central auditory system. In: Oertel, D., Popper, A.N., Fay, R.R. (Eds.), Integrative Functions in the Mammalian Auditory Pathway. Springer, New York, pp. 238–318.

Chang, T.R., Chung, P.C., Poon, P.W.F., Chen, E.L., Chiu, T.W., 2003. Responses of central auditory neurons modeled with finite impulse response (FIR) neural networks. Comput. Methods Programs Biomed. 74, 151–164.

Chang, T.R., Chung, P.C., Chiu, T.W., Poon, P.W., 2005. A new method for adjusting neural response jitter in the STRF obtained by spike-trigger averaging. Biosystems 79, 213–222.

Chang, T.R., Poon, P.W.F., Chiu, T.W., Chung, P.C., 2010a. Should neuronal spikes be given equal weighting in the generation of spectral temporal receptive field? J. Physiol. Paris 104, 215–222.

Chang, T.R., Sun, X., Poon, P.W.F., 2010b. Fine frequency-modulation trigger features of midbrain auditory neurons extracted by a progressive thresholding method. Chin. J. Physiol. 53, 430–438.

Chiu, T.W., Poon, P.W., 2007. Multiple-band trigger features of midbrain auditory neurons revealed in composite spectro-temporal receptive fields. Chin. J. Physiol. 50, 105–112.

Christianson, G.B., Sahani, M., Linden, J.F., 2008. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. J. Neurosci. 28, 446–455.

deCharms, R.C., Blake, D.T., Merzenich, M.M., 1998. Optimizing sound features for cortical neurons. Science 280, 1439–1443.

Depireux, D.A., Simon, J.Z., Klein, D.J., Shamma, S.A., 2001. Spectrotemporal response field characterization with dynamic ripples in ferret primary auditory cortex. J. Neurophysiol. 85, 1220–1234.

Eggermont, J.J., 1994. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. Hear. Res. 74, 51–66.

Eggermont, J.J., 2011. Context dependence of spectro-temporal receptive fields with implications for neural coding. Hear. Res. 271, 123–132.

Eggermont, J.J., Aertsen, A.M., Johannesma, P.I., 1983. Prediction of the responses of auditory neurons in the midbrain of the grass frog based on the spectrotemporal receptive field. Hear. Res. 10, 191–202.

Englitz, B., Ahrens, M., Tolnai, S., Rubsamen, R., Sahani, M., Jost, J., 2010. Multilinear models of single cell responses in the medial nucleus of the trapezoid body. Netw. Comput. Neural Syst. 21, 91–124.

Escabí, M.A., Read, H.L., 2005. Neural mechanisms for spectral analysis in the auditory midbrain, thalamus, and cortex. Int. Rev. Neurobiol. 70, 207–252.

Escabí, M.A., Schreiner, C.E., 2002. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. J. Neurosci. 22, 4114–4431.

Felsheim, C., Ostwald, J., 1996. Responses to exponential frequency modulations in the rat inferior colliculus. Hear. Res. 98, 137–151.
Fuzessery, Z.M., Hall, J.C., 1996. Role of GABA in shaping frequency tuning and creating FM sweep selectivity in the inferior colliculus. J. Neurophysiol. 76, 1059–1073.
Gittelman, J.X., Pollak, G.D., 2011. It's about time: how input timing is used and not used to create emergent properties in the auditory system. J. Neurosci. 16, 2576–2583.
Gittelman, J.X., Li, N., Pollak, G.D., 2009. Mechanisms underlying directional selectivity for frequency-modulated sweeps in the inferior colliculus revealed by in vivo whole-cell recordings. J. Neurosci. 29, 13030–13041.
Gollisch, T., 2006. Estimating receptive fields in the presence of spike-time jitter. Network 17, 103–129.
Gourevitch, B., Norena, A., Shaw, G., Eggermont, J.J., 2009. Spectrotemporal receptive fields in anesthetized cat primary auditory cortex are context dependent. Cereb. Cortex 19, 1448–1461.
Heil, P., Rajan, R., Irvine, D.R., 1992. Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I: effects of variation of stimulus parameters. Hear. Res. 63, 108–113.
Hermes, D.J., Aertsen, A.M., Johannesma, P.I., Eggermont, J.J., 1981. Spectro-temporal characteristics of single units in the auditory midbrain of the lightly anaesthetised grass frog (Rana temporaria L) investigated with noise stimuli. Hear. Res. 5, 147–178.
Kanwal, J.S., Rauschecker, J.P., 2007. Auditory cortex of bats and primates: managing species-specific calls for social communication. Front. Biosci. 12, 4621–4640.
Kao, M.C., Poon, P.W., Sun, X., 1997. Modeling of the response of midbrain auditory neurons in the rat to their vocalization sounds based on FM sensitivities. Biosystems 40, 103–119.
Kim, P.J., Young, E.D., 1994. Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers. J. Acoust. Soc. Am. 95, 410–422.
Klein, D.J., Depireux, D.A., Simon, J.Z., Shamma, S.A., 2000. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. J. Comput. Neurosci. 9, 85–111.
Lesica, N.A., Grothe, B., 2008. Dynamic spectrotemporal feature selectivity in the auditory midbrain. J. Neurosci. 28, 5412–5421.
Linden, J.F., Liu, R.C., Sahani, M., Schreiner, C.E., Merzenich, M.M., 2003. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. J. Neurophysiol. 90, 2660–2675.
Machens, C.K., Wehr, M.S., Zador, A.M., 2004. Linearity of cortical receptive fields measured with natural sounds. J. Neurosci. 24, 1089–1100.
Miller, L.M., Escabı´, M.A., Read, H.L., Schreiner, C.E., 2002. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. 87, 516–527.
Poon, P.W., Chiu, T.W., 2000. Similarities of FM and AM receptive space of single units at the auditory midbrain. Biosystems 58, 229–237.
Poon, P.W., Yu, P.P., 2000. Spectro-temporal receptive fields of midbrain auditory neurons in the rat obtained with frequency modulated stimulation. Neurosci. Lett. 289, 9–12.
Poon, P.W., Chen, X., Hwang, J.C., 1991. Basic determinants for FM responses in the inferior colliculus of rats. Exp. Brain Res. 83, 598–606.
Poon, P.W., Chen, X., Cheung, Y.M., 1992. Differences in FM response correlate with morphology of neurons in the rat inferior colliculus. Exp. Brain Res. 91, 94–104.
Qin, L., Wang, J., Sato, Y., 2008. Heterogeneous neuronal responses to frequency-modulated tones in the primary auditory cortex of awake cats. J. Neurophysiol. 100, 1622–1634.
Qiu, A., Schreiner, C.E., Escabí, M.A., 2003. Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. J. Neurophysiol. 2003 (90), 456–476.
Rees, A., Malmierca, M.S., 2006. Processing of dynamic spectral properties of sounds. Int. Rev. Neurobiol. 70, 299–330.
Rees, A., Møller, A.R., 1983. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. Hear. Res. 10, 301–330.
Reiss, L.A.J., Bandyopadhyay, S., Young, E.D., 2007. Effects of stimulus spectral contrast on receptive fields of dorsal cochlear nucleus neurons. J. Neurophysiol. 98, 2133–2143.
Rodríguez, F.A., Chen, C., Read, H.L., Escabí, M.A., 2010. Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. J. Neurosci. 30, 15969–15980.
Sivaramakrishnan, S., Oliver, D.L., 2001. Distinct K currents result in physiologically distinct cell types in the inferior colliculus of the rat. J. Neurosci. 21, 2861–2877.
Suta, D., Popelar, J., Syka, J., 2008. Coding of communication calls in the subcortical and cortical structures of the auditory system. Physiol. Res. 57 (Suppl. 3), S149–S159.
Theunissen, F.E., Sen, K., Doupe, A.J., 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J. Neurosci. 20, 2315–2331.
Valentine, P.A., Eggermont, J.J., 2004. Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex. Hear. Res. 196, 119–133.
Versnel, H., Zwiers, M.P., van Opstal, A.J., 2009. Spectrotemporal response properties of inferior colliculus neurons in alert monkey. J. Neurosci. 29, 9725–9739.
Whitfield, C., Evans, E.F., 1965. Responses of auditory cortical neurons to stimuli of changing frequency. J. Neurophysiol. 28, 655–672.
Ye, C.Q., Poo, M.M., Dan, Y., Zhang, X.H., 2010. Synaptic mechanisms of direction selectivity in primary auditory cortex. J. Neurosci. 30, 1861–1868.
Yeshurun, Y., Wollberg, Z., Dyn, N., 1989. Prediction of linear and non-linear responses of MGB neurons by system identification methods. Bull. Math. Biol. 51, 337–346.
Young, E.D., 2010. Level and spectrum. In: Rees, A., Palmer, A.R. (Eds.), The Oxford Handbook of Auditory Science: The Auditory Brain. Oxford Univ. Press, Oxford, pp. 93–124.
Young, E.D., Calhoun, B.M., 2005. Nonlinear modeling of auditory-nerve rate responses to wideband stimuli. J. Neurophysiol. 94, 4441–4454.
Zhang, L.I., Tan, A.Y., Schreiner, C.E., Merzenich, M.M., 2003. Topography and synaptic shaping of direction selectivity in primary auditory cortex. Nature 424, 201–205.