

Available online at www.sciencedirect.com
www.elsevier.com/locate/brainres

Research Report

Modeling complex responses of FM-sensitive cells in the auditory midbrain using a committee machine

T.R. Chang^a, T.W. Chiu^b, X. Sun^c, Paul W.F. Poon^{d,*}^aDepartment of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan^bDepartment of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan^cDivision of Life Sciences, East China Normal University, Shanghai, China^dDepartment of Physiology, Medical College, National Cheng Kung University, Tainan, Taiwan

ARTICLE INFO

Article history:

Accepted 30 April 2013

Available online 9 May 2013

ABSTRACT

Frequency modulation (FM) is an important building block of complex sounds that include speech signals. Exploring the neural mechanisms of FM coding with computer modeling could help understand how speech sounds are processed in the brain. Here, we modeled the single unit responses of auditory neurons recorded from the midbrain of anesthetized rats. These neurons displayed spectral temporal receptive fields (STRFs) that had multiple-trigger features, and were more complex than those with single-trigger features. Their responses have not been modeled satisfactorily with simple artificial neural networks, unlike neurons with simple-trigger features. To improve model performance, here we tested an approach with the committee machine. For a given neuron, the peri-stimulus time histogram (PSTH) was first generated in response to a repeated random FM tone, and peaks in the PSTH were segregated into groups based on the similarity of their pre-spike FM trigger features. Each group was then modeled using an artificial neural network with simple architecture, and, when necessary, by increasing the number of neurons in the hidden layer. After initial training, the artificial neural networks with their optimized weighting coefficients were pooled into a committee machine for training. Finally, the model performance was tested by prediction of the response of the same cell to a novel FM tone. The results showed improvement over simple artificial neural networks, supporting that trigger-feature-based modeling can be extended to cells with complex responses.

This article is part of a Special Issue entitled Neural Coding 2012.

This article is part of a Special Issue entitled Neural Coding 2012.

© 2013 Elsevier B.V. All rights reserved.

*Corresponding author. Fax: +886 6236 2780.

E-mail addresses: trchang@mail.stust.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).

1. Introduction

Time-varying signal of frequency modulation (FM) is an important building block of communication signals in both animals and humans (Lindblom and Studdert-Kennedy, 1967; Zeng et al., 2005; for review, see Kanwal and Rauschecker (2007)). Electrophysiological studies in animals showed that central auditory neurons are selectively sensitive to FM tones (Atencio et al., 2007; Brown and Harrison, 2009; Eggermont, 1994; Heil et al., 1992; Poon et al., 1991; Qin et al., 2008; Nelson et al., 1966; Rees and Møller, 1983; Whitfields and Evans, 1965; for review, see Suta et al. (2008)). The auditory 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; for review, see Rees and Malmierca (2005)). These FM-sensitive neurons respond mainly to tones of rapidly varying frequency, but not to tones of fixed frequencies. The neural mechanisms underlying FM coding remain elusive as critical studies require challenging techniques like in vivo whole cell patch clamping and the sample size is often limited (Gittelman et al., 2009; O'Neill and Brimijoin, 2002; Ye et al., 2010; Zhang et al., 2003). One approach to understand the neural mechanisms of FM coding is the computational modeling of spike responses to sound. This approach involves modeling the input-output relationship for a cell using one sound (probe), and then assessing the model performance using another sound (test). In modeling responses to complex sounds, results are more satisfactory at the lower centers, and less so at the higher centers (Kim and Young, 1994; Lesica and Grothe, 2008; Reiss et al., 2007; Theunissen et al., 2000). Such discrepancy in model performance was explained partly by the nonlinear properties of central circuits (Ahrens et al., 2008; Bar-Yosef et al., 2002; Christianson et al., 2008; Escabi and Schreiner, 2002; Young and Calhoun, 2005).

One common modeling strategy makes use of the spectro-temporal receptive field (STRF; Aertsen and Johannesma, 1981; Depireux et al., 2001; Eggermont et al., 1983; Hermes et al., 1981; Klein et al., 2000; Miller et al., 2002; Qiu et al., 2003; for review, see Yu and Young (2010)). The STRF is used to represent on the time–frequency plane the response probability of central auditory neurons to complex sounds. Typically, a probe tone of a randomly varying frequency is used to evoke spike responses (deCharms et al., 1998; Poon and Yu, 2000; Qiu et al., 2003; Theunissen et al., 2000). Averaging out the random sound energy preceding each spike in the spectro-temporal plane generates the STRF. Band-like concentrations of energy appear in the STRF of FM-sensitive cells. These bands are found at pre-spike intervals that are consistent with the neural transmission time measured from the auditory periphery to where the cell is recorded. These band-like structures represent the presumed ‘trigger features’ that determine the spike responses. In the STRFs of auditory neurons, a variety of trigger features have been reported (e.g., a flat orientation representing pure tone sensitivity; or a band displaying either a rising slope representing a modulation from low to high frequency, or a falling slope, a modulation from high to low frequency; Atencio et al., 2007; Chiu and Poon, 2007). The exact pattern of trigger features also

depends on the kind of sound used for stimulation, with a choice ranging from random tones to naturally-occurring sounds (Escabi and Schreiner, 2002; Poon and Yu, 2000; Theunissen et al., 2000; Valentine and Eggermont, 2004).

In a previous study (Chang et al., 2012) we have modeled the FM responses of auditory midbrain neurons based on trigger features derived from their STRFs. The results are satisfactory provided that (a) the neurons have simple receptive fields (i.e., a single trigger feature in the STRF), and (b) the spike responses are brief (i.e., phasic as opposed to sustained). Although about 80% of the FM-sensitive neurons in the midbrain satisfied the above criteria (Chiu and Poon, 2007), the remaining 20% displaying complex receptive fields (i.e., multiple trigger features in the STRF) yielded poor results with our model. Here, we hypothesize that neurons with complex receptive fields can be modeled using a composite neural network known as the committee machine. This approach assumes that individual trigger features are related to spike responses in a grossly linear fashion.

2. Results

A total of five neurons with complex receptive fields were analyzed. The detailed results of one (unit 43-2-5) are presented together with the key results of other neurons.

The STRFs of all neurons displayed more than one bands indicating the presence of complex or multiple trigger features (Fig. 1). The time profiles of PSTH peaks typically showed diverse shapes (single- or multiple-peaks, Fig. 2). The multiple-peaks appearance was consistent with a sustained response that typically dominated the PSTH. In contrast, the phasic responses were less common for these neurons. The diverse shapes in the PSTH responses suggested the presence of more than one type of FM sensitivities. The predicted results with finite impulse response neural networks (FIRNNs) model of simple architecture (1-1-1) were not satisfactory; even though the approximate time of response occurrence was grossly correct (Fig. 3). For a given cell showing complex responses to FM, the majority of the multiple-peak responses typically dominated the model output, and the minority of single peaks was not predicted or basically ignored by the model. Increasing the number of hidden neurons extended the model output to include the minor group of single peaks. However, the shape-fidelity of multiple peaks remained poorly modeled (e.g., even with as many as six neurons in the hidden layer) (Fig. 4, unit 100-3-2).

Comparing FM time profiles preceding the response and grouping them based on their similarity index had segregated PSTH peaks successfully into two or more groups (Fig. 2). Once grouped, each of them was modeled separately with its own FIRNN. The significant improvement, in terms of percentage of hit and the fidelity of multiple peaks, was observed only after the FIRNNs were pooled into a committee machine for modeling (Figs. 3 and 4).

Similar improvement of model responses was observed in four other neurons; Fig. 1 shows their STRFs, Fig. 2, their PSTH peak-groupings, and Fig. 4 shows the results before and after the separation of PSTH response-peaks. For the five neurons we analyzed, the improvement in response prediction with

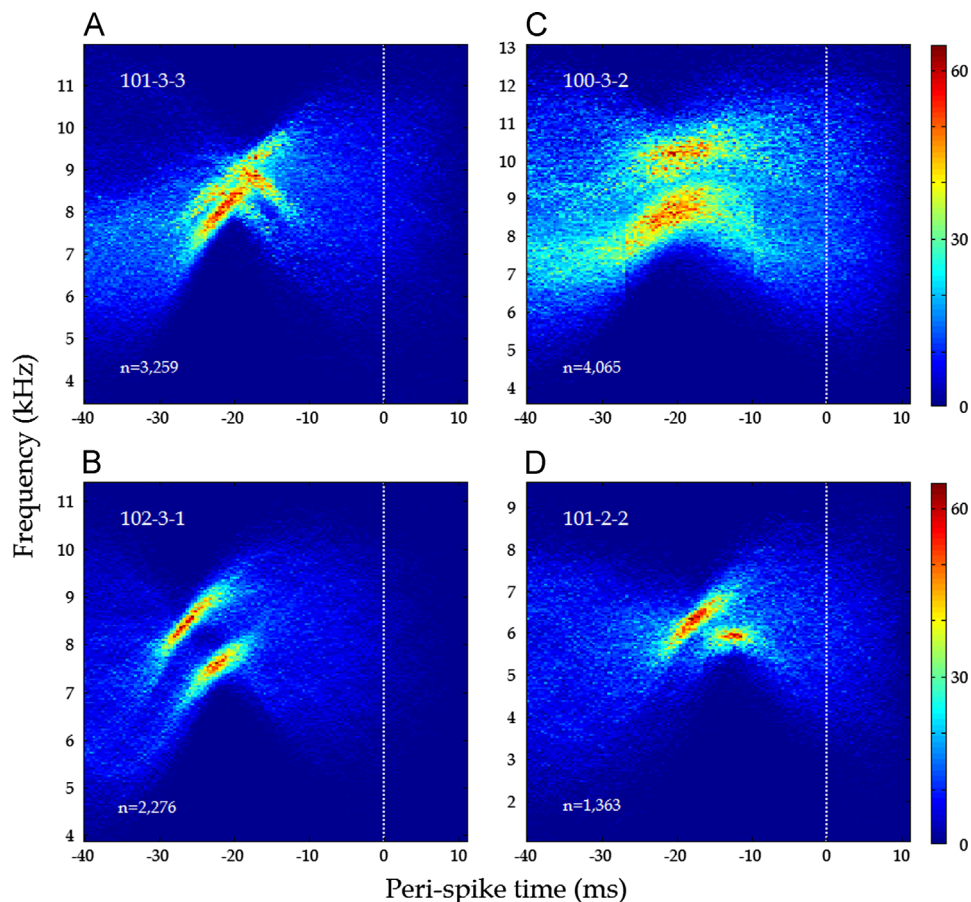


Fig. 1 – STRFs of four FM-sensitive neurons showing multiple trigger features after extraction with the procedures of spike-trigger averaging, progressive thresholding and de-jittering (see text for details). In this and other plots of STRFs (Fig. 6), the number of time-profiles overlapped at each pixel is color-coded to show the concentrated areas of trigger features. Note the orientations of the red bands representing different frequency modulations (e.g., rising FM, or modulation from low to high frequency; or falling FM, the reverse). The dashed line marks the time of spike occurrence, and n , the number of stimulus tracings in the plot.

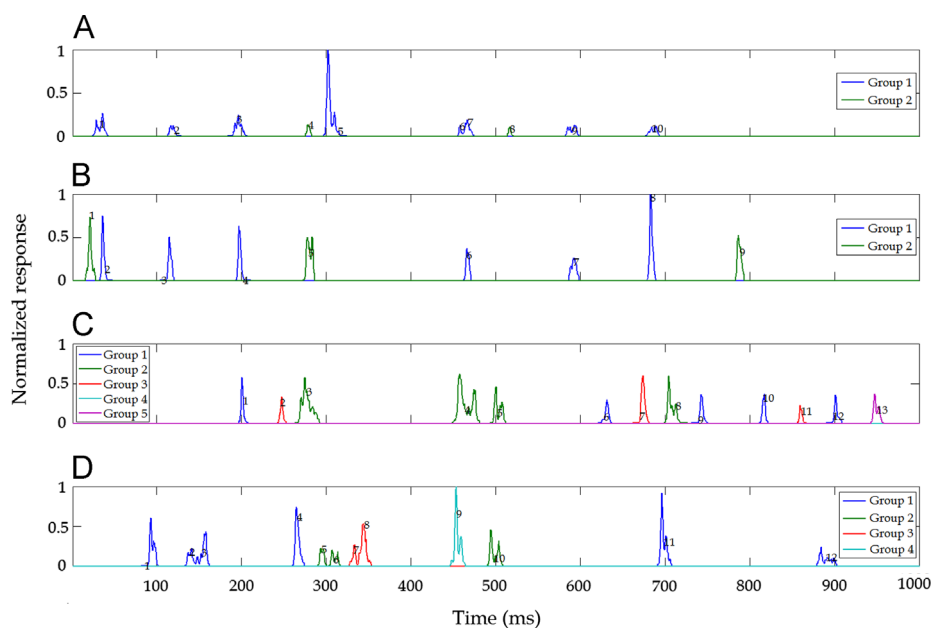


Fig. 2 – PSTHs of the four FM-sensitive neurons (same units as in Fig. 1) showing the grouping of response peaks according to the similarity of pre-spike trigger features. The groups of peaks (ranged from 2 to 5) so separated are shown in different colors and individual peaks numbered. See also Fig. 7D. Note in this and other plots of PSTH (Figs. 3, 4, and 7) the maximal response in each 2-s long data is normalized to 1. (A) unit 101-3-3, (B) unit 102-3-1, (C) unit 100-3-2 and (D) unit 101-2-2.

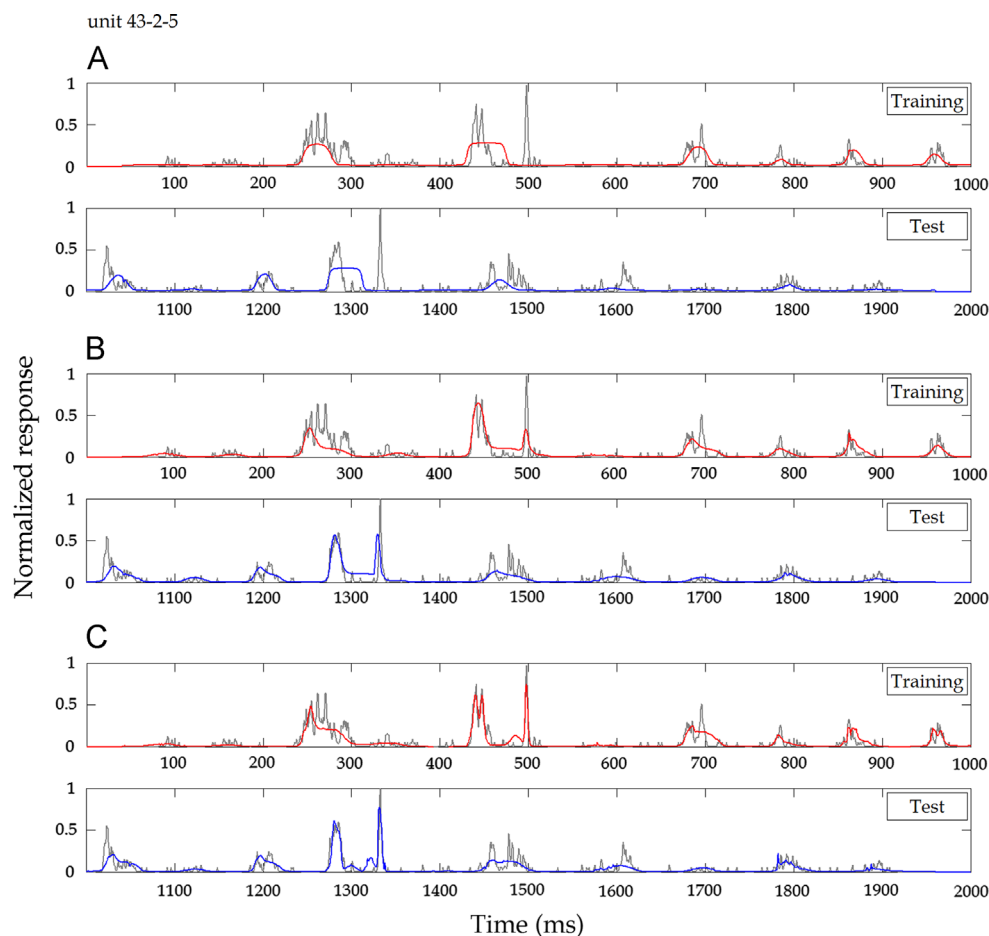


Fig. 3 – Representative results of modeling (unit 43-2-5). (A) with a simple FIRNN (architecture 1-1-1, representing one input neuron, one hidden neuron, one output neuron); (B) with a FIRNN of more hidden neurons, architecture 1-3-1 (one input neuron, three hidden neurons, one output neuron); and (C) with a committee machine of architecture 3-1-1 (three input neurons, one hidden neuron, one output neuron). The comparison is based on the same total number of artificial neurons used in (B) and (C) (i.e., $1+3+1=3+1+1$). Note the phasic responses at time 500, 1340 ms are better modeled after peak-grouping (see also Fig. 7D). Gray tracings represent the actual response; red tracings, the trained output; blue tracings, the predicted (test) output. Half of the PSTH was used for training (upper panels with red tracings) and the remaining half for testing (lower panels with blue tracings).

the committee machine over either the simple or extended FIRNN was statistically significant even for this small sample of five neurons (Wilcoxon rank sum test, $p < 0.01$).

3. Discussion

Our principal finding is that the FM-responses of midbrain auditory neurons were modeled more satisfactorily after segregating the spike responses into different groups based on their pre-spike trigger features, followed by a committee machine that had individual settings of model parameters. The advantage of applying different sets of model parameters is consistent with FM detections that involve different transmission delays and different time window of integration even for the same cell. Results supported the importance of trigger features in modeling FM responses, despite their complexity in STRF. For the same neurons, the poorer performance without PSTH peak separation is probably not surprising considering the basic principle of FIRNN modeling (i.e., to simulate the

output of diverse response peaks by optimizing only one set of model parameters). Evidently, FIRNNs of simple architecture (e.g., 1-1-1) worked well for neurons displaying simple STRFs since the response peaks likely shared a common spike-generating mechanism. However, their performance dropped in the case of neurons with complex STRFs. By carefully grouping the response peaks, and by optimizing the time window and delay time, the committee machine outperformed the single FIRNN, even when the total numbers of neurons in the two models were equal. It is also quite remarkable that given datasets as short as 1 s (and often with limited number of response peaks in the PSTH), our present model was able to predict relatively well the response to a novel stimulus. In particular, tone-1 and tone-2 were different in their frequency contents (tone-1 being higher and tone-2 lower in frequency contents). The purpose of tone-1 was to determine if the cell has a complex receptive field. If yes, tone-2 was used to generate data for training and testing of the model. We speculate that given a longer training dataset, the grouping of response peaks would be more precise and

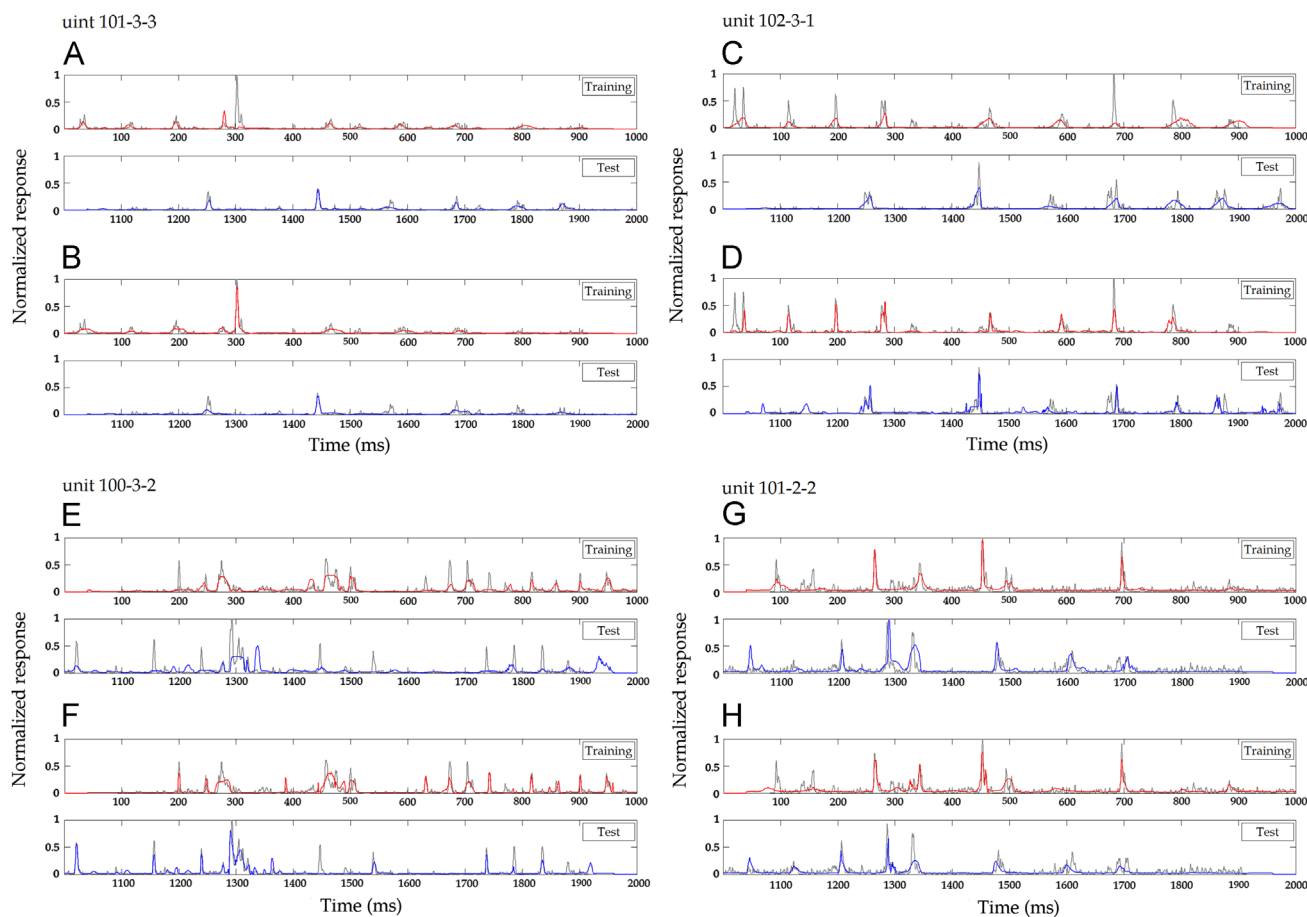


Fig. 4 – Results from another four neurons shown before and after peak-grouping (same units in Fig. 2). Captions are otherwise similar to that of Fig. 3. (A) 1-3-1, (B) 3-1-1, (C) 1-3-1, (D) 3-1-1, (E) 1-6-1, (F) 6-1-1, (G) 1-5-1 and (H) 5-1-1.

the performance of the model would be better. Since longer datasets will have more chance of training the model with minor trigger features that are not present in short datasets. Furthermore, increasing the group number, obtaining better parameter values will also improve model performance.

The multiple features shown in the STRF are used to demonstrate that the cell has a complex receptive field. The multiple groups of PSTH peaks obtained from the same cell should be consistent with the complex receptive field, if the same sound stimuli were used. However, in this study, STRF was derived from a fast FM, and PSTH from a slow FM. Hence the two stimuli had different FM features. Despite a difference in the FM stimuli for generating the STRF and PSTH, our approach does not depend on using the same sound stimuli. In fact, on an FM cell, given only the prior knowledge of a complex receptive field, our model would perform satisfactorily. This suggested that our model will work for cells with complex receptive fields in general, and it does not rely on the knowledge of the precise trigger feature.

A number of studies, both empirical and theoretical, have demonstrated nonlinear behaviors of the central circuitry in response to complex sounds (Christianson et al., 2008; Yu and Young, 2010). While our model does not rely heavily on non-linearity, this does not mean that non-linearity is not important at all. Our results simply showed that, at the level of the

midbrain, auditory neurons can still be considered to behave relatively linearly, provided that the stimulus is a random FM tone of a single frequency. Such stimulation with a single tone is apparently more manageable by our model. Should the FM stimulus be replaced by FM tones of multiple frequencies (e.g., like formants in speech signals), nonlinearity likely becomes more important. That is because the response of auditory neurons to two-tone stimulation is known not to be fully predicted by their responses to a single tone. The phenomenon of two-tone inhibition (i.e., the response to one tone is inhibited when another tone slightly off its characteristic frequency is presented simultaneously), has been known to occur at the level of the auditory nerve fibers which reflects the cochlear nonlinearity (Sachs and Kiang, 1968). In parallel with our committee machine approach, other investigators have used a multilinear model to improve the prediction of responses of auditory neurons to complex sounds (Ahrens et al., 2008). However the comparison of the results between their approach and ours is not straightforward as they did not use a single tone for stimulation. These multilinear or composite models are superior to simple neural networks.

Although the sampled neurons had characteristic frequencies covering a wide range from 500 to 50 kHz, those with complex receptive fields fell within the mid-frequency range that also overlapped with the most sensitive part of the

rat's audiogram. What biological implications this finding has is not clear.

In summary, our present approach, based on multiple trigger features in the STRF, improves model performance to FM tones including cells with complex receptive fields. It would be most interesting to extend the present approach by modeling the responses to naturally-occurring complex sounds.

4. Experimental procedures

Detailed experimental protocols have been reported before (Chiu and Poon, 2000; Chang et al., 2010a, 2012), and they were approved by the Animal Ethics Committee of NCKU Medical College, Taiwan. In brief, using KCl-filled glass micropipettes, the single unit responses of auditory midbrain neurons to FM sounds were recorded extracellularly in anesthetized rats (Sprague Dawley strain, 250–350 g body weight, urethane, 2 g/kg, intra-peritoneal). Two sets of FM sounds were presented: (a) random FM (tone-1) that was different across trials, and (b) random FM (tone-2) that was identical across trials. The FM stimuli were obtained by low-pass-filtering white noise (125 Hz for tone-1, 25 Hz for tone-2) and the output signal was used to control the instantaneous frequency of a sine wave generator. During data collection, stimuli were delivered in the free-field to the animal, at ~ 30 dB suprathreshold, ± 1 octave across the best frequency of the cell. For each cell, FM tone-1 was used to generate the STRF, the first part of FM tone-2 (1 s long) for model training, and the remaining part (1 s long) for model testing. The peaks in the PSTH are the responses to be modeled (Fig. 5).

To generate an STRF, the instantaneous frequency time-profiles of the FM stimulus within a 40 ms pre-spike time window were added according to the conventional procedure. Specifically, on the spectro-temporal plane formed by a minimum of 126×201 pixels, the count of all the stimulus frequency time-profile passing through each pixel was registered to show the concentration of peri-spike sound energy (Poon and Yu, 2000). To better reveal the trigger features in the STRF, we applied a preprocessing procedure called 'progressive thresholding' (Chang et al., 2010b) followed by 'de-jittering' (Chang et al., 2005). In brief, for each STRF, pixels with counts greater than a preset level (e.g., 55% of maximum) were identified as a 'supra-threshold area' and its boundary extracted. In the case of detecting more than one area, the one larger in size was processed first. Those pre-spike time tracings of modulating waveforms passing through the supra-threshold area were removed from the population and grouped under one trigger feature. This procedure was repeated iteratively at progressively higher threshold levels (+5% for each iteration) until all tracings were pooled into separate groups. For each group of tracings, a procedure of de-jittering was applied to correct for the response jitter that was transferred erroneously to the stimulus tracings due to the limitations of spike-trigger averaging. This procedure involved first matching individual modulating waveforms to their mean at a systematically varied time window to reach an optimal variance time profile. At this optimal window length, the delay of each tracing was then adjusted systematically to minimize the disparity with respect to the population mean. The purpose was to reveal individual trigger features in the blurred raw STRF. After processing, individual trigger features typically appear in the

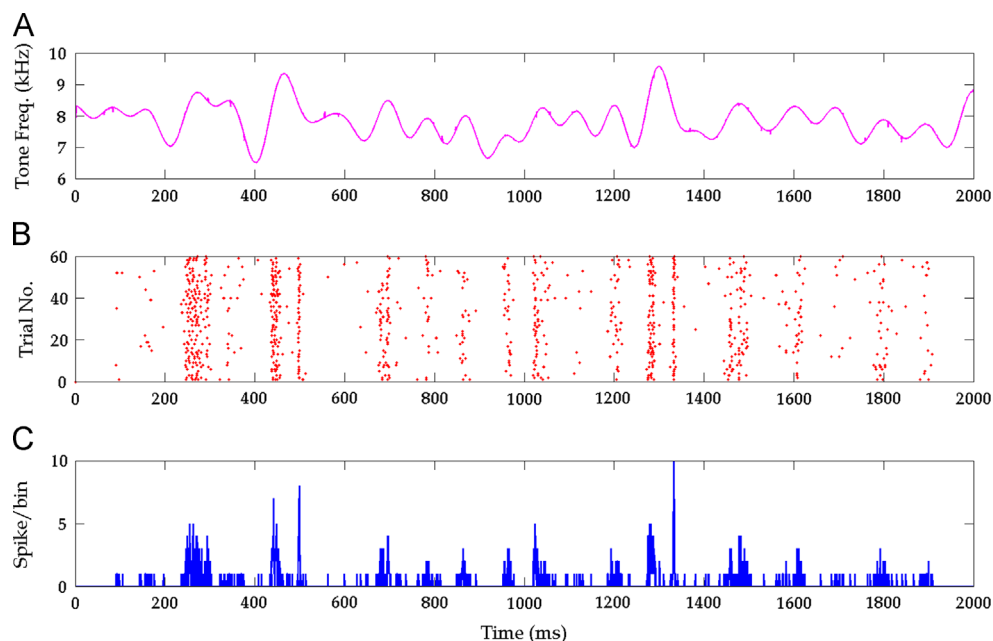


Fig. 5 – Responses of a representative FM neuron (unit 43-2-5): (A) instantaneous frequency time-profile of a random FM stimulus (tone-2), (B) the corresponding spike responses to 60 repeated stimulus trials, shown in dot-raster (each dot represents an action potential), and (C) spike responses in (B) plotted as peri-stimulus time histogram (PSTH), before the Gaussian filtering. Note the peaks represent presumed activation by the trigger features in the stimulus. Half of the 2-s dataset is used for model training, and the remaining half for testing.

form of a single band-like structure (i.e., simple receptive field), or as multiple band-like structures (complex receptive field; Fig. 6). We analyzed only those neurons displaying complex receptive fields. Due to the complex response pattern,

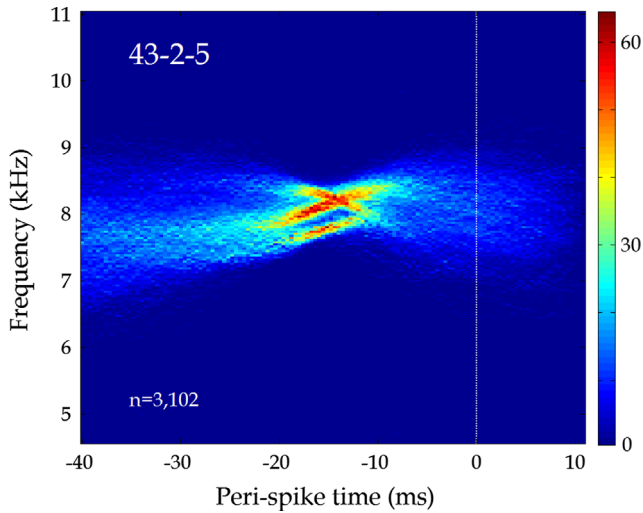


Fig. 6 – Representative STRF showing complex receptive field (same unit as in Fig. 5). Captions are otherwise similar to that of Fig. 1.

peaks in the PSTH that had exceeded a baseline level (10% of maximum response) were extracted and grouped according to their similarity of pre-spike FM time profiles based on an algorithm of minimal disparity (for details please see [Supplementary material 1](#)). For each group of peaks, the temporal information of the trigger feature (i.e., its start and end times; and the delay between the end of the trigger feature and the onset of spike response, i.e., the central transmission delay) were optimized before setting the parameters of the artificial neural networks. The optimizing procedure involved a systematic scanning across the pre-spike time range for the optimal combination of time window and delay that gave the maximal model performance (Fig. 7). The optimal combination, according to our previous study ([Chang et al., 2012](#)), was taken near the tip of a triangular area forming a plateau-like structure, or at the peak of an island-like structure in the scan-result plots (arrows in Fig. 7). To assess model performance, we used the percentage of overlap in the PSTH as the index: here we compute the area of positive hit ('intercept') and divide it by the total area ('union') of actual and predicted responses. Hence we consider only the areas of PSTH with spike responses and ignore those without ([Chang et al., 2012](#)). This performance index represents what the artificial neural network was trained to maximize.

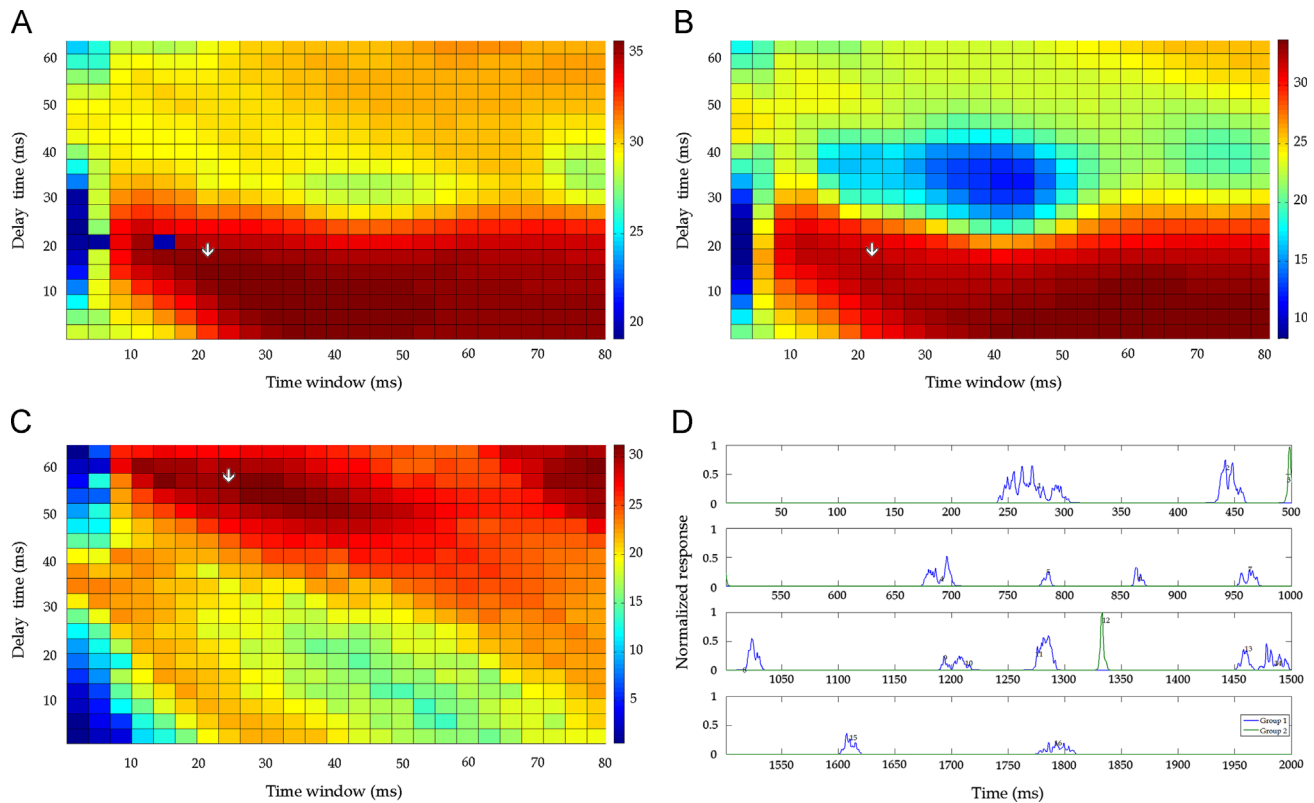


Fig. 7 – Model performance plotted as the results of systematic varying the time-window width and delay time (same unit as in [Figs. 5 and 6](#)). The model performance is color-coded to represent its percentage overlap with PSTH responsive areas (warm colors, better performance). Training performances based on all peaks in the PSTH (A), and the separated group-1 peaks (B), and group-2 peaks (C). The two groups of peaks so separated are shown in the PSTH with different colors and numbered for reference (D). Arrows in (A)–(C) indicate the optimal combinations of time window and delay time for each set of peaks. For the simplicity of illustration, results here were based on the analysis of 2-s data segments. For the calculation of model performance, please see text.

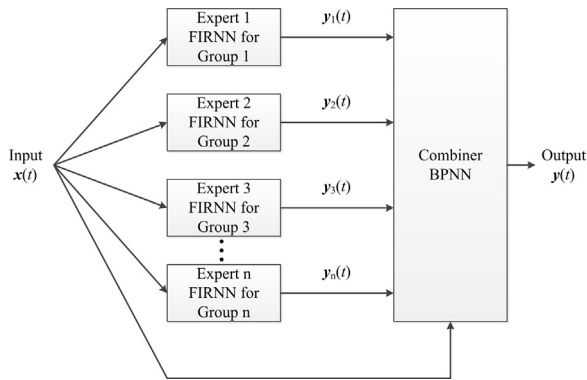


Fig. 8 – Schematic architecture of a committee machine that contains a number of finite impulse response neural networks (FIRNNs or expert systems). Each expert or FIRNN was trained to model an individual peak-group in the PSTH. The combiner of FIRNNs is a back propagation neural network (BPNN).

In essence, the artificial neural network was trained to generate the spike response probability based on the instantaneous frequency time-profiles of the FM stimulus using the optimized pre-spike time window and central transmission delay. Because of response grouping, different settings of the model parameters were allowed for different groups. The number of neurons in the hidden layer was free to increase from the default 1 to a maximum of 5 in order to improve the fidelity of modeling the PSTH response time profile. To simulate synaptic activation of the cell and to facilitate modeling the response probability profile a multi-scale Gaussian function was convolved with the PSTH. In more technical terms, a finite impulse response neural network (FIRNN) which modeled synaptic interactions as FIR linear filters was constructed in the form of an autoregressive time series (for technical details please see [Chang et al. \(2012\)](#)). After successful modeling of individual groups, the FIRNNs were pooled into a committee machine, which further incorporated the input FM stimulus time profile that was required for adjusting the relative weighting of the FIRNNs ([Fig. 8](#)). A committee machine has a parallel architecture that produces an output from combining the results of individual experts ([Dietterich, 2000](#)). In this study, we used an ensemble-based committee machine with a number of empirical formulae. To include the contribution of individual formulae, the output of each weighted member (FIRNN) and the input FM stimulus time profile are computed according to the weights (coefficients) of the weighted average. The optimal combination of weights for prediction is processed using a back propagation neural network (BPNN; for more technical details on committee machine and BPNN please see [Supplementary material 2](#)).

Finally, the performance of the committee machine was accessed with an FM tone (the novel part of tone-2).

Acknowledgments

We thank Dr. Iain Bruce for commenting on the manuscript; supported by grants from the National Science Council (nos. 101-2221-E-218-010-MY2, 99-2320-B-006-020-MY3, and 100-2923-B-006-001-MY3).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.brainres.2013.04.058>.

REFERENCES

- Aertsen, A.M., Johannesma, P.I., 1981. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol. Cybern.* 42, 133–143.
- 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.
- 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.
- Chang, T.R., Chung, P.C., Chiu, T.W., Poon, P.W.F., 2005. A new method for adjusting neural response jitter in the STRF obtained by spike-trigger averaging. *Biosystems* 79, 213–222.
- Chang, T.R., Chiu, T.W., Chung, P.C., Poon, P.W.F., 2010a. Should spikes be treated with equal weightings in the generation of spectro-temporal receptive fields?. *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.
- Chang, T.R., Chiu, T.W., Sun, X., Poon, P.W.F., 2012. Modeling frequency modulated responses of midbrain auditory neurons based on trigger features and artificial neural networks. *Brain Res.* 1434, 90–101.
- Chiu, T.W., Poon, P.W.F., 2000. Basic response determinants for amplitude modulation of single neurons in the auditory midbrain. *Exp. Brain Res.* 134, 237–245.
- Chiu, T.W., Poon, P.W.F., 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., Blacke, D.T., Merzenich, M.M., 1998. Optimizing sound features for cortical neurons. *Science* 280, 1439–1442.
- 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.
- Dietterich, T.G., 2000. Ensemble methods in machine learning. *Multiple Classifier Syst.* 1857, 1–15.
- 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.
- 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.

- Escabi, M.A., Schreiner, C.E., 2002. Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J. Neurosci.* 22, 4114–4131.
- Felsheim, C., Ostwald, J., 1996. Responses to exponential frequency modulations in the rat inferior colliculus. *Hear. Res.* 98, 137–151.
- 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.
- 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 (*Ranatemporaria 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.
- 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.
- 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.
- Lesica, N.A., Grothe, B., 2008. Dynamic spectrotemporal feature selectivity in the auditory midbrain. *J. Neurosci.* 28, 5412–5421.
- Lindblom, B., Studdert-Kennedy, M., 1967. On the role of formant transitions in vowel recognition. *J. Acoust. Soc. Am.* 42, 830–843.
- Miller, L.M., Escabi, M.A., Read, H.L., Schreiner, C.E., 2002. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J. Neurophysiol.* 87, 516–527.
- Nelson, P.G., Erulkar, S.D., Bryan, J.S., 1966. Responses of units of the inferior colliculus to time-varying acoustic stimuli. *J. Neurophysiol.* 29, 834–860.
- O'Neill, W.E., Brimijoin, W.O., 2002. Directional selectivity for FM sweeps in the supragenicular nucleus of the mustached bat medial geniculate body. *J. Neurophysiol.* 88, 172–187.
- Poon, P.W.F., Chen, X., Hwang, J.C., 1991. Basic determinants for FM responses in the inferior colliculus of rats. *Exp. Brain Res.* 83, 598–601.
- Poon, P.W.F., 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.
- Poon, P.W.F., 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.
- 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., Escabi, M.A., 2003. Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition. *J. Neurophysiol.* 90, 456–476.
- Rees, A., Malmierca, M.S., 2005. 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.
- Sachs, M.B., Kiang, N.Y., 1968. Two-tone inhibition in auditory-nerve fibers. *J. Acoust. Soc. Am.* 43, 1120–1128.
- 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.
- Whitfield, I.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.
- Young, E.D., Calhoun, B.M., 2005. Nonlinear modeling of auditory-nerve rate responses to wideband stimuli. *J. Neurophysiol.* 94, 4441–4454.
- Yu, J.J., Young, E.D., 2010. Linear and non-linear pathways of spectral information transmission in the cochlear nucleus. *Proc. Natl. Acad. Sci.* 97, 11780–11786.
- Zeng, F.G., Nie, K., Stickney, G.S., Kong, Y.Y., Vongphoe, M., Bhargava, A., Wei, C., Cao, K., 2005. Speech recognition with amplitude and frequency modulations. *Proc. Natl. Acad. Sci.* 102, 2293–2298.
- 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.