Reviewer #1 (Advances the Field):  
  
See comments to authors.  
  
Reviewer #1 :  
  
The paper by Thomas at all is elegant in its simplicity. They scan a participant once to encode the tonotopic population RFs and then they scan the participant again this time stimulating them with a drawn out version of two melodic songs (because of the slow temporal resolution of fMRI). Then they decode the tone frequencies presented. I find a number of appealing aspects to this paper, thus it is worth seeing published, as I note. However, my main concern is that although the paper used tone sequences, it is not really about decoding sequences but tones (the sequencing may not have mattered if it was random tone transitions being decoded).  
  
Positives:  
It's a relatively straight forward approach, once the method is developed and shown to work by a group like the one here, which is not trivial. Also, I would not have guessed a priori that it would work as well as it did in individual subjects. They are also unlikely to be find the same voxel again on the second session given the between scanning session variability (distortions, etc.) that even the best co-registration software may struggle to compensate for. In a nutshell, I'm surprised that it works so well in individual subjects and as the authors note the approach can be more flexibly applied than other classifier approaches. Likely the approach does not depend on the same voxel but an adjacent voxel may do as well, which is very nice to see.  
  
Those of us using BOLD pattern classification may view this as being overly simple tonotopic decoding, but I do find the paper appealing and the approach is likely to be useful to the community.   
  
Concern:  
The paper is more about decoding tone frequencies than sequences and that needs to be clearer throughout: Even though the authors use melodic sequences during scanning, I had originally thought from the title that they were decoding a rapid sequence with the pattern of voxel activity. The paper is still interesting but I'm not sure whether it mattered if the stimulus was a drawn out melodic sequence or a random set of tones. The first order Markov chain used to generate simulated sequences without the melodic sequence is used as a reference to evaluate the significance of the effects. If these had been used as stimuli during fMRI I'm sure the tones could have been decoded as well since the algorithm is not by design sensitive to the transitional probabilities between the notes and how they evolve over time as the participant is listening. This is worth considering more carefully in the paper or as a topic for future study.   
  
Relatedly, on line 208 the authors suggest that they use the first order Markov chain to match first and second order statistical properties, but a second order Markov process is needed to impact on second order statistical regularities. Were there second order transitional regularities that the approach was able to decode that I missed?  
  
Minor: I would have been nice to hear sound demonstrations of the stimuli and decoded versions from the brain.  
  
  
Reviewer #2 (Rationale for Significance/Novelty Rating):  
  
The study is incremental to recent previous work by the same group. There are novel elements but are rather limited (see review)  
  
Reviewer #2 :  
  
This article by Thomas et al. uses fMRI and a population receptive field (pRF) approach to examine whether and to what extent it is possible to reconstruct a sequence of 2-s tones (songs) from fMRI response patterns. To a large extent, the article is a confirmation of previous work by the same group (Thomas et al, Neuroimage 2015), which showed that a population receptive approach similar to that used in the visual domain can be applied to study the representation of frequency and tonotopic maps in auditory cortex. As in this previous study, stimuli and analysis are limited to one stimulus dimension (frequency). The novel aspect included here is the prf-based reconstruction of (slow) tone sequences. While this finding is presented as a major breakthrough, it is basically a confirmation that frequency maps obtained with random tone sequences are similar to those obtained used tone sequences forming a melody. This may be interesting, but certainly not unexpected, given the  
literature reviewed in the introduction and showing that - for primary areas - similar tonotopic maps are obtained with different stimuli and paradigms. Beyond this, the article could provide additional information on the representation of sound frequency in the auditory cortex, but current experiments and analyses are rather limited. I detail below my concerns and specific comments.  
  
Title is too general: "...reconstructing auditory sequences..." should be "...reconstructing tone sequences..."  
  
Abstract: The last sentence is very generic and not very useful; replace with a statement on the significance of the actual results of the study.  
  
Introduction: I have found a bit confusing that in a paragraph it is stated that different paradigms lead to consistent tonotopic maps, while in another following paragraph it is stated that the travelling wave approach may introduce biases. Please clarify.  
  
Introduction/Methods/Results: The authors write: "Here, we present a method for investigating the accuracy of our voxel-wise frequency tuning curves...". However, the analyses and results presented do so only very indirectly. In fact, if I understood correctly, the analysis of reconstruction accuracy is based on a multivoxel score based on the fmincon\* matlab function. Thus the obtained score refers to the all auditory cortex and tells very little about the accuracy of a single voxel.

No analysis is presented to illustrate which voxels actually contributed to the reconstruction. Where were they located? Are those with narrowest tuning? Is the reconstruction obtained A1 voxels and R voxels comparable? These are interesting analyses that have not been included so far.  
  
\*Line 180: correct "fminconv" with "fmincon" - Provide more detail on the function parameters and how this function was used.  
  
Methods: It is a strong limitation of the current study that only two tone sequences (songs) were used to test the predictive value of the prf model. Using a larger variety of test sequences would have provided larger possibilities in the analysis and allowed addressing interesting questions. E.g. one could have studied the accuracy of frequency reconstruction at different octaves, or for predictable vs unpredictable melodies. Also, one could have manipulated the frequency range, repetition rate, loudness etc. While it is clear that some of this can be done in further studies, including at least one of the manipulations in the current study would have increased its appeal.   
  
Results:   
Figure 2B arrows for the frequency gradients are drawn manually. Justify or remove.  
  
I find puzzling in the reconstruction results that while the correlation values are comparable across subjects (Figure 3) the residual errors (Table 1) are hugely variable across subjects. This means that even with a very coarse reconstruction of the frequency (Subject 2) leads to high identification scores. Could you please elaborate and explain this?  
  
As mentioned, it would be interesting to perform and report the analyses for separate regions of the AC (e.g. A1 vs R) or for separate group of voxels (e.g. narrow vs broadly tuned). This would be an important addition to the paper.  
  
I understand that this type of studies do not require a large number of subjects. However, if the goal is to assess the accuracy of fMRI frequency maps and this is quantified using the residual errors reported in table 1, then a larger number of subjects is needed (given the current large variability across subjects).  
  
Discussion  
  
It would be important to remark the current limitations of the proposed approach. Especially in comparison to some of the reconstruction studies using visual stimuli, the stimuli used here were very simple and varied only in one dimension (frequency). Also, the frequency content of the stimuli used for the reconstruction analysis is constant for a relatively large time period (two seconds) and that this is a rare case for natural sounds. Thus, moving to the reconstruction of more realistic sound would require large extensions to the proposed methods.  
  
  
  
  
  
------------------------------------------------------------------------  
  
The Journal of Neuroscience is a member of the Neuroscience Peer Review Consortium, an alliance of neuroscience journals that have agreed to share manuscript reviews at the author's request. Its goals are to speed the publication of research reports and reduce the burden on authors and peer reviewers. The most appropriate manuscripts for this process are those that have been judged as being technically sound and requiring modest revision, but which were not considered to be appropriate for the Journal's broad audience. Experience has shown that papers that require substantial revision due to technical problems or other issues probably should be submitted de novo to the second journal.  
  
For information about the Consortium, the participating journals and instructions on how to forward reviews, please visit:  
  
<http://nprc.incf.org/authors/information-for-authors>