Modeling vocalization with ECoG cortical activity recorded during vocal production in the macaque monkey

Makoto Fukushima, Richard C. Saunders, Naotaka Fujii, Bruno B. Averbeck, Mortimer Mishkin

Abstract— **Vocal production is an example of controlled motor behavior with high temporal precision. Previous studies have decoded auditory evoked cortical activity while monkeys listened to vocalization sounds. On the other hand, there have been few attempts at decoding motor cortical activity during vocal production. Here we recorded cortical activity during vocal production in the macaque with a chronically implanted electrocorticographic (ECoG) electrode array. The array detected robust activity in motor cortex during vocal production. We used a nonlinear dynamical model of the vocal organ to reduce the dimensionality of 'Coo' calls produced by the monkey. We then used linear regression to evaluate the information in motor cortical activity for this reduced representation of calls. This simple linear model accounted for circa 65% of the variance in the reduced sound representations, supporting the feasibility of using the dynamical model of the vocal organ for decoding motor cortical activity during vocal production.**

I. INTRODUCTION

The auditory cortex processes sounds it receives with high temporal precision to achieve auditory perception of complex natural sounds such as vocalizations[1]-[4]. Previous studies have shown robust decoding of auditory cortical activity recorded from ECoG electrode arrays when subjects listen to vocalizations of humans and monkeys [5]-[7]. On the other hand, there have been relatively few attempts to decode cortical activity during vocal production. In monkeys, neural activity associated with motor commands has been observed in multiple cortical areas during vocal production [8]-[10]. Recently a nonlinear dynamical model of the vocal organs was used to synthesize the song of zebra finch with its complex spectral temporal structures [11]-[13]. This model is based on a flapping mechanism originally suggested for human vocal folds [14], [15], that can reduce complex spectrotemporal sound features to two time-varying parameters: "air pressure" and "muscle tension". It is biologically plausible that the brain controls such parameters to produce complex sounds. Moreover, reducing the high dimensionality of emitted sounds could be advantageous in predicting or reconstructing vocalizations from neural signals recorded from a limited number of recording sites, as in the application of a brain-machine interface for speech production [16].

Here we trained a rhesus monkey to vocalize for water rewards, and then chronically implanted an ECoG electrode array to record field potentials from numerous cortical areas during vocal production. This array robustly detected electrophysiological cortical activity during vocal production in the macaque monkey. We then derived a reduced representation of the vocalizations by fitting the nonlinear dynamical model of the vocal organ to each call produced by the monkey. Finally, we evaluated how much of the variance in the model parameters was explained by a linear regression model from neuronal activity recording in the motor cortex.

II. METHOD

A. Subjects

An adult male rhesus monkey (Macaca mulatta) was used in the current study. All procedures and animal care were conducted in accordance with the Institute of Laboratory Animal Resources Guide for the Care and Use of Laboratory Animals. All experimental procedures were approved by the National Institute of Mental Health Animal Care and Use Committee.

B. Electrophysiological and sound recording during vocal production

During the experiment, the monkey was placed in a sound-attenuating booth (Bioacoustics Instruments). The monkey vocalized to obtain water reward with its head fixed. We monitored the monkey's behavioral state through a video camera and microphone connected to a PC, and we also recorded eye movements during the experiment. The monkey's vocalization was recorded with a directional microphone (Audio Technica) with a sampling rate of 25kHz. The auditory evoked potentials from the 256 channels of the ECoG array were band-passed between 2 and 500 Hz, digitally sampled at 1500 Hz, and stored on hard-disk drives by a PZ2 preamplifier and the RZ2 base station (Tucker Davis Technologies).

C. ECoG electrode

The ECoG electrode array consists of 256 recording sites for bipolar recording at 128 locations (Fig. 1): the electrodes on the medial wall of the left hemisphere (26 sites) were designed to cover the medial frontal cortex, cingulate cortex, and supplementary motor area. The electrodes on the lateral surface (212 sites) were designed to cover the frontal, temporal, parietal, and occipital lobes. The electrodes in the lateral sulcus (18 sites) were designed to cover 2 cm of the

This research is supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health, and Department of Health and Human Services. M. Fukushima, R.C. Saunders, B.B. Averbeck, and M. Mishkin are with the National Institute of Mental Health, Bethesda, MD 20892 USA (phone: 301-443-7458; fax: 301-402-0046 e-mail: makoto.fukushima@nih.gov). N. Fujii is with RIKEN Brain Science Institute, Saitama, Japan.

caudorostral cortical surface on the ventral bank of the lateral sulcus. Each recording site was a circular disk with 0.8 mm diameter and the distance between two sites in a bipolar pair was 1.8mm. The mean impedance values ranged from 140kΩ (at 10Hz) to 0.9 kΩ (at 10kHz).

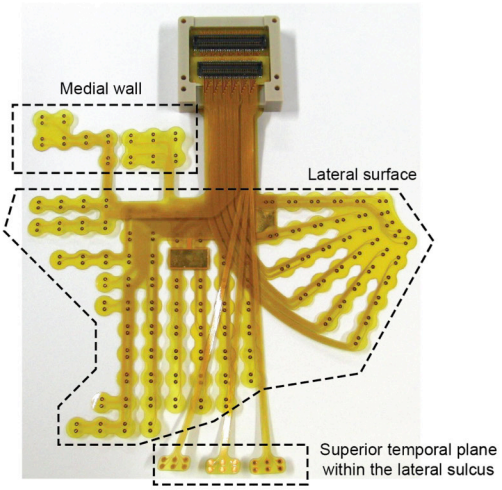

Figure 1. The 256 channel ECoG array. Spatial layout of all 256 electrodes. The three sections outlined by the dashed lines were designed for the medial wall (26 electrodes), the lateral surface (212 electrodes), and the supratemporal plane of the lateral sulcus (18 electrodes)

D. Nonlinear dynamical model of the vocal organ

We used the following nonlinear dynamical model of the vocal organ [11], [12] to derive a reduced representation of monkey vocalizations.

$$dx(t)/dt = y(t) \tag{1}$$

$$dy(t)/dt = -\alpha\gamma^2 - \beta\gamma^2 x - \gamma^2 x^3 - \gamma x^2 y + \gamma^2 x^2 - \gamma xy \tag{2}$$

For each monkey 'Coo' call (Fig. 2a), we estimated two time-varying parameters (α:"pressure" and β:"tension"), following a similar procedure used for zebra finch songs [17]. The variable γ is a time invariant constant. $x(t)$ is the labial position of the model, which was used to calculate the sound pressure at the vocal tract, $P_i(t)$, using:

$$P_i(t) = x(t) - rP_i(t-\tau) \tag{3}$$

We then calculated the output sound pressure from the vocal tract, $P_t(t)$, by:

$$P_t(t) = (1-r)P_i(t-\tau) \tag{4}$$

This $P_t(t)$ was used to fit vocalization sounds recorded from the monkey. Here r =-0.9, and τ = 0.11 ms [12]. Each monkey call was decomposed into consecutive 20 ms segments. The consecutive segments were shifted by 2 ms (i.e. 18ms overlap). For each 20 ms segment of monkey call, we chose the optimal pair of α and β that provided the most similar power spectrum distribution of the segment. We did a grid search to find this optimal pair (100x100; $0<\alpha<0.2$, $0<\beta<0.4$). Previous studies of zebra finch songs [12], [18] used γ=~24000, but we found that γ=5000 provided a good fit for the monkey 'Coo' calls. After α and β were estimated from segmented vocalizations, the time courses of α and β were smoothed by low-pass filtering at 20Hz (Fig. 2b). Then we numerically solved (1) and (2), with these time-varying α and β. This numerical solution provided reconstructed monkey calls (Fig. 2a, right). The differential equation was solved with an ODE solver ('ode45') in MATLAB (MathWorks).

E. Linear regression analysis

We used the following liner regression model to quantify how much of the variance in the parameters of the vocal organ model could be explained by cortical activity in the motor cortex.

$$p(t) = h(0) + \sum_{k=1}^{M} h(k)r(t-k) \tag{5}$$

where $p(t)$ stands for either α or β obtained by fitting the dynamical model to a monkey call (Fig. 2b). M specifies the number of data points included in the regression model. $h(k)$ is the regression coefficient. $r(t)$ is the amplitude of high-gamma power. This is obtained by band-pass filtering the field potential with a Butterworth filter (70-200Hz, 4th

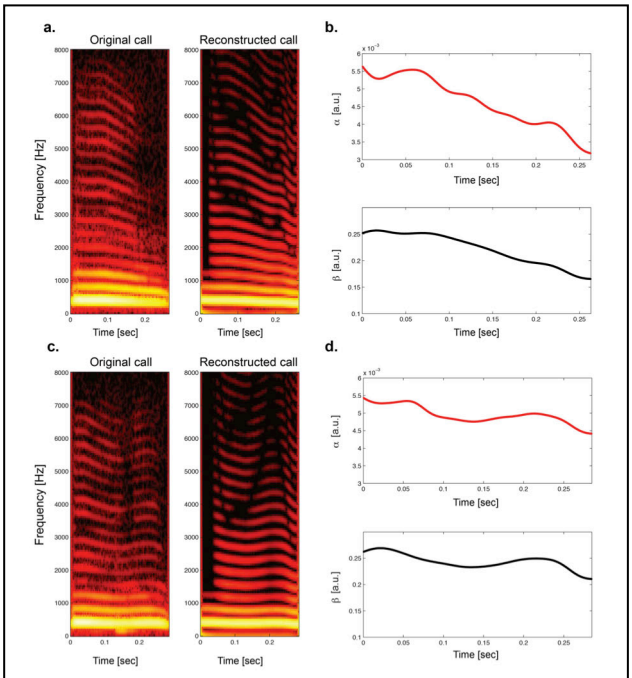

Figure 2. Two example monkey calls reconstructed by the dynamical model. a. an example monkey call (left), and the reconstructed call (right). **b.** two parameters α and β that produced the reconstructed call in a. **c.** another expample call (left), and the reconstructed call (right). **b.** two parameters produced the reconstructed call in **c**.

order for low-pass, 7th order for high-pass), and then the amplitude was obtained by taking the absolute value of the band-pass waveform. We then smoothed it by low-pass filtering at 10Hz (4th order Butterworth filter). Followed by down-sampling the data at a 250Hz sampling rate. For all filtering processes, we achieved a zero-phase shift by processing the data in both forward and reverse directions in time.

III. Results

A. Reduced representation of recorded 'Coo' calls

As noted, we focused our analysis on harmonically structured 'Coo' calls produced by the monkey and fit the dynamical model to these calls. This resulted in a reconstructed call (Fig. 2a and 2c right) that was very similar to the original call (Fig. 2a and 2c, left). Although these two exmaple calls are different in the amount of frequency modulatation over time, the model reprodurced the original calls well. This demostrates the utility of the model in prodcuing reduced representations for 'Coo' calls with different spectrotemporal profiles. We reconstructed 30 'Coo' calls produced by the monkey, and used associated parameters α and β (Fig. 2b) in the following analysis.

B. Cortical activation during vocal production

The ECoG array recorded robust cortical field potentials during vocal production. The normalized spectrogram from a bipolar recording site in the motor cortex showed increases in power from the baseline period in the gamma frequency range (50–200 Hz; Fig. 3), starting from around 600 ms before the onset of vocalization. On the other hand, the low frequency power (4-30Hz) was decreased below the baseline, approximately 1 second before the onset of vocalization (Fig. 3).

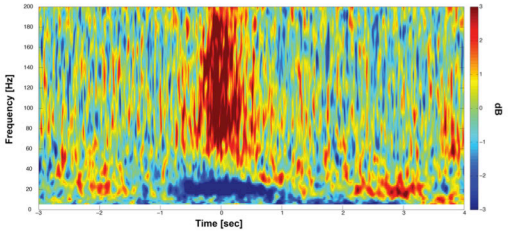

Figure 3. Spectrogram of motor cortical response during vocal production. This is the average of 30 calls. The call onset is aligned to 0 in the time axis. The spectrogram is normalized by the basiline period acivtiy which is 4 seconds before the onset of the call.

C. Motor cortical activity explains variability in the vocal parameters

Motor cortical activity during vocal production sould be correlated with vocal parameters if the cortical activity plays a role in controlling the vocal organs. Thus, we used the linear regression model to quantify how much of the variance in the estimated vocal parameters can be explained by the motor cortical acitivity. We used the high-gamma power amplitude from one bipolar site around the motor cortex to predict the vocal parameters (see Methods). We calucated the linear regression for different *M* values that specify the number of time points from the high gamma amplitude. We found M=887 (=3.55 sec at 250 Hz sampling rate) to be the largest value that was significant at $p<0.01$. In this case, the high-gamma power explained 65.2% of the variance in the α paramter ($p=0.0096$, $F_{(887,\ 567)}=1.1969$) and 66.5% of the variance in the β paramter ($p=0.0009$, $F_{(887,\ 567)}=1.2700$). Accordindgly the regression models for each of those two parameters exhibited reasonable fits to the data (Fig. 4).

IV. Discussion

In the current study, we recorded cortical activity with an ECoG electrode array during vocal production in a macaque monkey. We then derived a reduced representation of monkey calls using a nonlinear dynamical model of the vocal organ. This allowed us to correlate motor cortical activity with estimated vocal parameters using a linear regression model. The results showed a significant correlation between motor cortical activity and the vocal parameters.

This framework may be useful for decoding motor cortical activity with a nonlinear dynamical model during vocal production. There are several issues that need to be addressed to further this aim. We used the model of the vocal organ for songbirds to reduce the dimensionality of monkey vocalizations, and thus the model could further be improved by incorporating biophysical characteristics of the vocal organs in primates. Under the constraints of the current model, some of the parameters could be chosen better. The current study optimized the variable γ, for example, only for 'Coo' calls, but a better parameter value may be chosen to fit the model to other types of monkey calls (e.g. grunt, scream, bark, etc.).

Decoding the cortical activity requires a statistical model to predict vocal parameters in independent data sets. In this case, the prediction may be difficult using the linear model. Thus, the next step for this analysis would be to predict the vocal parameters with cross-validated data sets, possibly with non-linear models of the type previously used in predicting motor commands [19]. Another possibility is to take advantage of simultaneous recording from multiple cortical areas. By combining multiple channels, the predictor performance might be substantially improved. This may require regularized methods for regression because of the high dimensionality of the data [7], [20].

Acknowledgment

We thank Matt Mullarkey and Alex Doyle for technical assistance. This study utilized the high-performance computational capabilities of the Helix Systems (http://helix.nih.gov) and the Biowulf Linux cluster (http://biowulf.nih.gov) at the National Institutes of Health, Bethesda, MD.

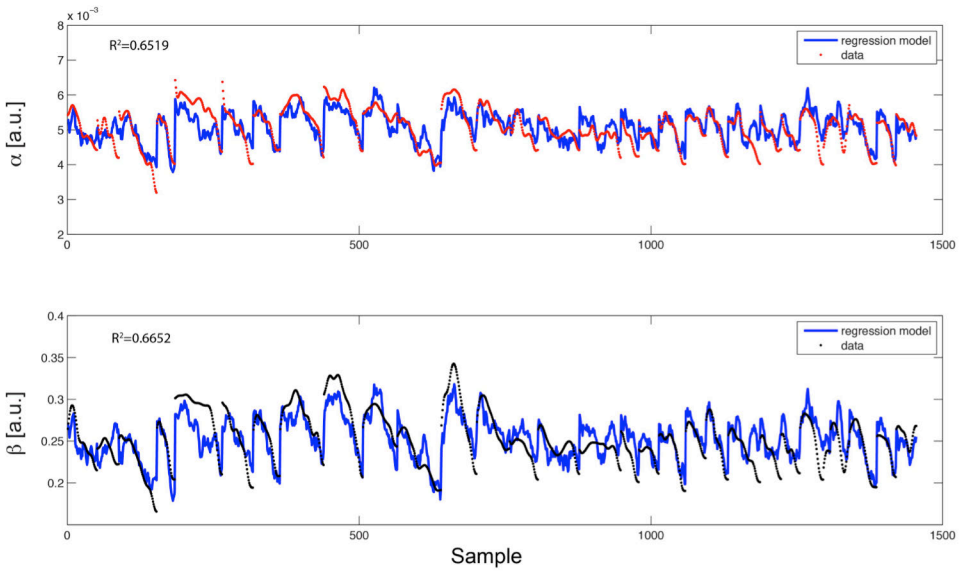

Figure 4. Linear regression model of the high-gamma amplitude from the motor cortex predicts the vocal parameters. The estimated vocal parameter was concatenated for all 30 calls. Upper panel: α paramter ("air pressure"). Lower panel : β parameter ("muscle tention"). The regression line is shown in blue, and the estimated parameter from calls plotted with dots.

REFERENCES

[1] Y. Kikuchi, B. Horwitz, and M. Mishkin, "Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane.," *J Neurosci*, vol. 30, no. 39, pp. 13021–13030, Sep. 2010.

[2] C. Perrodin, C. Kayser, N. K. Logothetis, and C. I. Petkov, "Voice Cells in the Primate Temporal Lobe," *Current Biology*, vol. 21, no. 16, pp. 1408–1415, Aug. 2011.

[3] X. Wang, "On cortical coding of vocal communication sounds in primates.," *P Natl Acad Sci Usa*, vol. 97, no. 22, pp. 11843–11849, Oct. 2000.

[4] E. Smith, S. Kellis, P. House, and B. Greger, "Decoding stimulus identity from multi-unit activity and local field potentials along the ventral auditory stream in the awake primate: implications for cortical neural prostheses," *J Neural Eng*, vol. 10, no. 1, p. 016010, Jan. 2013.

[5] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex.," *PLoS Biol*, vol. 10, no. 1, p. e1001251, Jan. 2012.

[6] E. F. Chang, J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight, "Categorical speech representation in human superior temporal gyrus," *Nat Neurosci*, vol. 13, no. 11, pp. 1428–1432, Oct. 2010.

[7] M. Fukushima, R. C. Saunders, D. A. Leopold, M. Mishkin, and B. B. Averbeck, "Differential coding of conspecific vocalizations in the ventral auditory cortical stream.," *J Neurosci*, vol. 34, no. 13, pp. 4665–4676, Mar. 2014.

[8] R. A. West and C. R. Larson, "Neurons of the anterior mesial cortex related to faciovocal activity in the awake monkey.," *J Neurophysiol*, vol. 74, no. 5, pp. 1856–1869, Nov. 1995.

[9] G. Coudé, P. F. Ferrari, F. Rodà, M. Maranesi, E. Borelli, V. Veroni, F. Monti, S. Rozzi, and L. Fogassi, "Neurons Controlling Voluntary Vocalization in the Macaque Ventral Premotor Cortex," *PLoS ONE*, vol. 6, no. 11, p. e26822, Nov. 2011.

[10] A. Nieder and S. R. Hage, "Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations," *Nature Communications*, vol. 4, pp. 1–11, Aug. 2013.

[11] A. Amador and G. B. Mindlin, "Beyond harmonic sounds in a simple model for birdsong production," *Chaos*, vol. 18, no. 4, p. 043123, 2008.

[12] J. D. Sitt, E. M. Arneodo, F. Goller, and G. B. Mindlin, "Physiologically driven avian vocal synthesizer," *Phys. Rev. E*, vol. 81, no. 3, p. 031927, Mar. 2010.

[13] A. Amador, Y. S. Perl, G. B. Mindlin, and D. Margoliash, "Elemental gesture dynamics are encoded by song premotor cortical neurons," *Nature*, pp. 1–7, Feb. 2013.

[14] I. R. Titze, "The physics of small-amplitude oscillation of the vocal folds.," *J Acoust Soc Am*, vol. 83, no. 4, pp. 1536–1552, Apr. 1988.

[15] E. M. Arneodo, Y. S. Perl, F. Goller, and G. B. Mindlin, "Prosthetic Avian Vocal Organ Controlled by a Freely Behaving Bird Based on a Low Dimensional Model of the Biomechanical Periphery," *PLoS Comput Biol*, vol. 8, no. 6, p. e1002546, Jun. 2012.

[16] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, R. Law, S. A. Siebert, J. L. Bartels, D. S. Andreasen, P. Ehirim, H. Mao, and P. R. Kennedy, "A Wireless Brain-Machine Interface for Real-Time Speech Synthesis," *PLoS ONE*, vol. 4, no. 12, p. e8218, Dec. 2009.

[17] J. Sitt, A. Amador, F. Goller, and G. Mindlin, "Dynamical origin of spectrally rich vocalizations in birdsong," *Phys. Rev. E*, vol. 78, no. 1, p. 011905, Jul. 2008.

[18] Y. S. Perl, E. M. Arneodo, A. Amador, F. Goller, and G. B. Mindlin, "Reconstruction of physiological instructions from Zebra finch song," *Phys. Rev. E*, vol. 84, no. 5, p. 051909, Nov. 2011.

[19] B. B. Averbeck, M. V. Chafee, D. A. Crowe, and A. P. Georgopoulos, "Parietal representation of hand velocity in a copy task," *J Neurophysiol*, vol. 93, no. 1, pp. 508–518, 2005.

[20] A. V. Cruz, N. Mallet, P. J. Magill, P. Brown, and B. B. Averbeck, "Effects of dopamine depletion on information flow between the subthalamic nucleus and external globus pallidus," *J Neurophysiol*, vol. 106, no. 4, pp. 2012–2023, Oct. 2011.