

Using Speech and Electroconvulsive Therapy to Map Human Auditory Cortex

Jeremy D.W. Greenlee, Roozbeh Behroozmand, Kirill V. Nourski, Hiroyuki Oya, Hiroto Kawasaki, Matthew A. Howard, III

Abstract— Much less is known about the organization of the human auditory cortex compared to non-human primate auditory cortices. In an effort to further investigate the response properties of human auditory cortex, we present preliminary findings from human subjects implanted with depth electrodes in Heschl’s gyrus (HG) as part of their neurosurgical treatment of epilepsy. Each subject had electroconvulsive therapy (ECoG) responses taken from medial and lateral HG in response to both speech and non-speech stimuli, including during speech production. Responses were somewhat variable across subjects, but posteromedial HG demonstrated frequency following responses to the stimuli in all subjects to some degree. Results and implications are discussed.

I. INTRODUCTION

The organization of the human auditory cortex is the least understood of all primate auditory cortices. Unlike non-human primates, exquisite anatomical tracer studies are not possible in humans and as a result the numbers and borders of the subregions comprising primary and higher order human auditory cortex are not known. While it is generally believed that humans utilize the same overall schema of non-human primate auditory cortex, namely a ‘core’ region, surrounded and reciprocally connected with a ‘belt’ region, and a ‘parabelt’ region which is reciprocally connected with the belt fields. Each of these three regions consists of multiple areas [1-6].

Functional neuroimaging is one experimental tool used to map auditory cortex [e.g. 7]. For example, fMRI provides excellent spatial resolution but temporal resolution inherent in BOLD responses does not allow effective comparisons of latency differences within auditory cortical regions. Conversely, scalp EEG is a technique which offers fine-grain temporal resolution (i.e. milliseconds) but limited spatial resolution due to intrinsic lowpass filtering properties of the skull and scalp.

Research supported by grants from the NIDCD.

J.D.W.G. Author is with the Department of Neurosurgery, University of Iowa, Iowa City, IA, 52242, USA (phone: 319-356-2771; fax: 319-353-6605; e-mail: jeremy-greenlee@uiowa.edu).

R.B. Author is with the Department of Neurosurgery, University of Iowa, Iowa City, IA, 52242, USA (e-mail: r-behroozmand@uiowa.edu).

K.V.N. Author is with the Department of Neurosurgery, University of Iowa, Iowa City, IA, 52242, USA (e-mail: kirill-nourski@uiowa.edu).

H.O. Author is with the Department of Neurosurgery, University of Iowa, Iowa City, IA, 52242, USA (e-mail: hiroyuki-oya@uiowa.edu).

H.K. Author is with the Department of Neurosurgery, University of Iowa, Iowa City, IA, 52242, USA (e-mail: hiroto-kawasaki@uiowa.edu).

M.A.H. Author, III is with the Department of Neurosurgery, University of Iowa, Iowa City, IA, 52242, USA (e-mail: matthew-howard@uiowa.edu).

Electroconvulsive therapy (ECoG) in neurosurgical patients is an effective basic research tool to further our understanding of human cortical physiology [8-16]. ECoG offers a combination of exquisite temporal resolution (milliseconds) with fine-grain anatomical resolution (millimeters). By simultaneously recording from multiple auditory regions during sound processing tasks, ECoG provides unique insights into human auditory cortical structure and function. This report describes the use of ECoG to better understand cortical processing of self-generated speech sounds compared to non-speech sounds.

II. METHODS

A. Subjects

Subjects were patient-volunteers undergoing the neurosurgical treatment of medically refractory epilepsy. These patients required implanted electrode arrays for ECoG to clinically identify and localize seizure foci. The same electrode arrays can be used for research studies such as this. All research protocols have been approved by the University of Iowa IRB, and all subjects gave informed consent to participate in research. Subjects did not incur additional medical risk by participating in this study.

Subjects in this report were a subset of patients from a larger cohort completing a variety of ECoG investigations. Inclusion criteria included the presence of a penetrating depth electrode array implanted along the long axis of Heschl’s gyrus (HG; [17]). All subjects underwent pre-operative detailed neuropsychological testing which confirmed the absence of speech or language impairment. Standard clinical audiometry confirmed normal hearing in all subjects. No anatomic lesions were found involving HG on imaging studies.

B. Tasks and stimuli

The goal of this study was to further understand the structure and function of human auditory cortex. In particular, we examined processing of speech and non-speech acoustical stimuli, as well as self-generated speech sounds and compared the resultant auditory cortical responses with those obtained during playback of the same self-generated speech stimuli. Because human speech has periodicity (e.g. voice fundamental frequency, F0), we compared these stimuli with non-speech stimuli containing periodicity (e.g. 100 Hz click trains). Speech production consisted of sustained vowel phonation (/a/). This vowel was chosen because it provides a relatively stable, uninterrupted F0 throughout the vocalization. Subjects

produced these utterances at their own pace and at normal conversational loudness, with their only instruction being to try to produce consistent loudness and vocal pitch from trial to trial. Approximately 50 such vocalizations were obtained for averaging suitable to ECoG analyses. All speech epochs were captured with a handheld microphone (Beta 87, Shure, Niles, IL), amplified (+10 dB, Mark of the Unicorn, Cambridge, MA), and recorded on a multichannel data acquisition system (System3, Tucker-Davis Technologies (TDT), Alachua, FL). Voice onsets were determined manually from the recorded sound waveform. Recorded vocalizations were played back to subjects via insert earphones (ER4, Etymotic, Elk Grove Village, IL) placed in custom-fit, vented ear molds. Gain of playback stimuli was adjusted such that the measured sound intensity was equal to that captured during speaking trials. We recognize that bone conduction is present during speaking and not during playback.

The non-speech acoustic stimuli for the study were click trains. These were a series of 100 Hz clicks presented from the TDT system to the earphones. Such clicks are useful to identify frequency following responses (FFR) within HG [15, 16, 19]. Clicks were digitally generated as equally-spaced rectangular pulses (0.2 ms duration) and were presented at a rate of 100 Hz (train duration 160 ms). Subjects were not required to respond to the clicks or given specific instructions to attend or not-attend to the stimuli.

C. ECoG electrodes and techniques

For this report, all ECoG data was obtained from depth electrodes placed into the cortex of HG. The details of our electrodes and implantation, recording, electrode localization, and data analysis techniques have been described [10-16]. In brief depth electrode arrays were placed into HG using stereotactic guidance (Stealth, Medtronic, Minneapolis, MN). Each array had at least 4 equally spaced low-impedance ('macro') ring contacts along the shaft of the electrode (Ad-Tech, Racine, WI). In addition, depth electrodes used in some subjects contained high-impedance microwire ('micro') contacts that extended off the side of the electrode shaft. The microcontacts were spaced from the medial extent of the electrode to the lateral surface of the superior temporal gyrus (STG). The position of each electrode contact in each subject was determined using a combination of high-resolution photographs taken during implantation and removal surgeries, post-operative CT scan, and pre- and post-operative thin-cut MRI scans. The estimated error in localization using these techniques does not exceed 2 mm.

D. Data analysis

ECoG data were digitized (2034.5Hz) and filtered (0.7 – 1000 Hz) by the TDT system online before resampling to 2000Hz offline. All data analysis was performed using custom scripts in MATLAB (Mathworks, Natick, MA). Individual trials for each dataset were inspected and trials containing artifacts or ictal activity were discarded prior to averaging. Time frequency responses (event-related band power; ERBP) were calculated using Morlet wavelet

decomposition. Power was determined relative to a pre-stimulus baseline period, typically -0.4 to -0.2 s prior to stimulus onset.

III. RESULTS

It has previously been reported on the lateral surface STG responses to vocalization and playback [9, 11]. Those reports found that high gamma (70-150 Hz) ERBP responses on STG were generally attenuated in amplitude during vocalization compared to playback at the majority of sites on STG. However, some STG sites demonstrated no difference in high gamma power between the 2 conditions, and an even smaller number of contacts showed increased high gamma amplitude during vocalization. We have not yet observed FFRs to voice on lateral STG.

In contrast to the lateral STG responses to self-vocalization, but consistent with previous reports [15,16, 19], the data we now present demonstrated that posteromedial HG showed strong FFRs to 100 Hz click trains. Figure 1 shows an example subject (L275), with recordings taken from a left sided HG electrode (Fig.1A). Fig.1B shows strong, narrow band ERBP activity centered at 100 Hz, beginning just after click onset and persisting for the duration of the click train. These responses were localized to the medial half of HG. Also observed in the same posteromedial HG contacts is a more broadband high gamma response of shorter duration than the narrow band FFR (Fig.1B, contacts L2-L4).

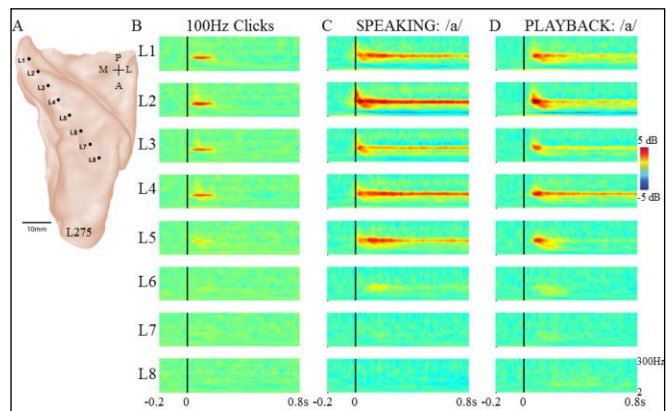


Figure 1. Responses from left HG in subject 275 (A) demonstrating strong narrow-band FFR to 100 Hz clicks (B), voice F0 during sustained vowel production (C) and playback of vowel production (D). Note shorter FFR latency during vocal production (C) compared to playback (D). All plots show 0.2 ms before stimulus onset, 0.8 ms after onset, 2-300 Hz, and +/- 5 dB.

This subject's HG response to self-vocalization was dramatically different compared to results other studies have shown for lateral STG responses to self-vocalization [9,11]. As shown in Fig.1C, the same posteromedial HG contacts that showed FFR to clicks showed a narrow-band ERBP response to self-vocalization. The subject's mean produced F0 was 124 Hz, and the peak magnitude of the narrowband high gamma ERBP during vocalization was centered at this same frequency. This FFR to F0 persisted throughout the vocalization. Like the speech production task, playback of

the same utterances also evoked FFR to F0 in the same posteromedial HG contacts (Fig.1D). Of interest is that 2 of the contacts (L3, L5) showed modulation of the FFR to F0. Namely, at these sites, the FFR during speaking was only transient during the playback condition while it persisted during speech production (contact L5) or diminished in amplitude (contact L3; Fig.1C vs Fig.1D). Like the responses to clicks, a more distributed high gamma ERBP response is evident in both speaking and playback conditions at the same contacts in addition to the FFR to F0.

A different example subject demonstrated a different HG response pattern. This subject (L173) also had left-sided HG electrode coverage with both macro- and microcontacts (Fig.2A). Like the previous subject, this subject demonstrated FFRs in posteromedial HG to 100 Hz click trains (Fig.2B). Click FFRs were seen only in the posteromedial third of HG. The FFRs were accompanied by shorter duration high gamma responses. Unlike the previous subject, this subject did not demonstrate FFR to F0 during either speaking or playback tasks (Figs.2C, D). There were prominent high gamma responses in both posteromedial and anterolateral portions of HG during both speaking (Fig.2C) and playback (Fig.2D). A notable difference between these 2 subjects is the F0, as L173's mean produced F0 was 282 Hz (vs 124Hz of subject L275 in Fig.1). Interestingly, it is known that the capacity of posteromedial HG to phase lock to click trains diminishes as click rates increase such that FFRs have not been observed for stimuli with rates near or above 200 Hz [16].

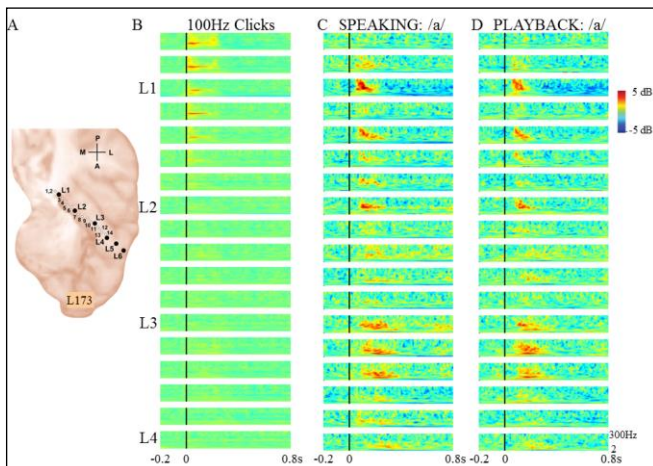


Figure 2. Responses from left HG in subject 173 (A) demonstrating strong narrow-band FFR to 100 Hz clicks (B). While large high gamma responses are seen on HG for both speech conditions, no FFRs to voice F0 occurred during either speaking (C) or playback of vowel production (D).

A third subject demonstrated yet another HG response pattern. This subject (L175) also had a left HG electrode (Fig.3A.) and his mean F0 produced was 114 Hz. Relatively weak FFRs were seen to 100 Hz click trains in posteromedial HG (Fig.3B). During speaking and playback conditions, FFRs to voice F0 (Figs.3C, D) were much stronger in these same HG contacts in addition to contacts medial and lateral to those that responded to clicks. In contrast to the subject illustrated in Fig.1, this subject's FFR to F0 was not

significantly modulated during speaking compared to playback (Fig.3C vs Fig.3D).

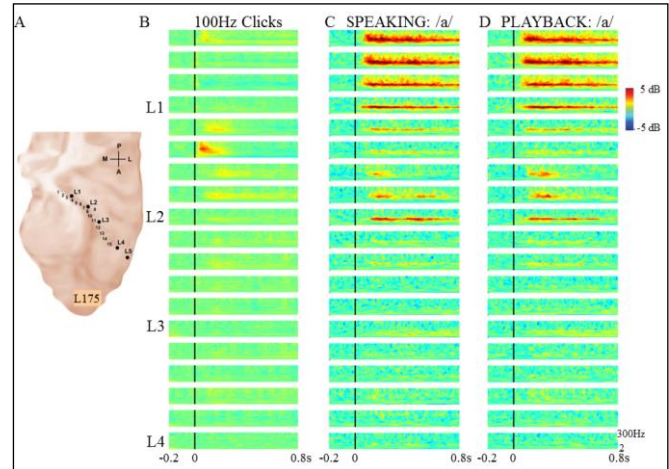


Figure 3. Responses from left HG in subject 175 (A) demonstrating weaker narrow-band FFR to 100 Hz clicks (B) but stronger and a larger FFR response area to voice F0 during both sustained vowel production (C) and playback of vowel production (D).

Although not a focus of this report, it is evident that response onset latency differences occur within HG to these stimuli. Specifically, each of these three subjects showed responses in posteromedial HG occurring with shorter onset latencies than sites just anterolaterally. This pattern was seen for both response to clicks as well as responses to speech production and playback. Previous reports have reported similar latency patterns in human HG [15,16].

IV. DISCUSSION

By utilizing both speech and non-speech sound stimuli, we have presented preliminary findings demonstrating examples of differential processing of sound within human HG. One subject (L275, Fig.1) demonstrated anatomic overlap of FFRs to both 100 Hz click trains and his mean produced F0 (124 Hz) during both speech production and playback conditions. A second subject (L173, Fig.2) with a much higher F0 (282 Hz) showed a FFR to clicks but did not show FFR to her F0 during either speech production or playback. A third example subject (L175, Fig.3) showed FFRs that were more robust for voice F0 than those to click trains in posteromedial HG. Taken together, the data suggest that FFRs to F0 may be a mechanism for vocal monitoring during vocal production for those subjects with F0's less than 200Hz. Those speakers with higher F0s may utilize different neural mechanisms (i.e. non-FFR) for vocal monitoring during speech production, such as high gamma responses.

The data show less modulation of higher frequency (FFR and high gamma) responses in posteromedial HG during speech production versus playback than the degrees of modulation seen on lateral STG [9-11,19]. This observation may reflect the different cortico-cortical connections of core auditory regions compared to belt or parabelt fields [e.g.20], and may be a manifestation of feedforward commands

influencing auditory cortical responses during speech production (e.g. [21]).

Clearly, more studies are needed to replicate and understand these exciting preliminary findings in a larger series of subjects. Several factors might contribute to the varied responses we report. For example, our electrode trajectories sample only a portion of HG, and HG has a complex gyral and sulcal pattern that can vary from subject to subject, along with variations in HG electrode placement. These different depth electrode trajectories may sample different subfields within HG that are known to exist. A larger number of subjects is required to better understand the field structure within HG and the respective response properties given the relatively sparse sampling provided by our depth electrodes.

Another important variable to consider, particularly for the auditory core region within HG, is the inherent differences in the energy delivered to the cochlea between speaking and playback conditions. Since bone conduction is absent during playback, the total stimulus energy is greater during speech production (i.e. air + bone conduction) compared to playback (air only). Core areas are particularly sensitive to stimulus intensity differences [22]. Yet, our multi-contact ECoG simultaneous recording paradigm provides relatively large numbers of recording electrodes such that each patient can serve as their own control where some contacts may show no difference in FFR or high gamma response intensity between conditions, some contacts show *greater* response intensity during speaking, and some contacts show *weaker* response intensity during speaking. In addition, it is not possible to effectively measure and control for degrees of attention paid to the tasks by each subject. Finally, the degrees of vocal (musical) training or experience were not assessed in these subjects, which might impact the precision of vocal production and F0 variability.

Future studies will need to further explore the range of voice F0 that HG FFR can track. Experiments employing repetitive stimuli of various types (click, tone, voice) and various frequencies will be needed to define the response properties of both medial and lateral HG, and examine how these properties compare with lateral STG.

V. CONCLUSION

This report demonstrates the use of speech and non-speech stimuli with high-resolution ECoG to identify differential responses within human HG. The response patterns we present highlight the ability of primary auditory cortex on posteromedial HG to follow the periodicity of these stimuli.

REFERENCES

- [1] Hackett TA, Stepniewska I, Kaas JH. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in Macaque monkeys. *J Comp Neurol* 495:475–95.
- [2] Hackett TA, Preuss TM, Kaas J. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197–222.
- [3] Hackett TA. (2008). Anatomical organization of the auditory cortex. *J Amer Acad Audiol* 19:774–9.

- [4] Hackett TA. (2011). Information flow in the auditory cortical network. *Hear Res* 271:133–46.
- [5] Kaas JH, Hackett TA. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA* 97:11793–99.
- [6] Romanski LM, Averbeck BB. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Ann Rev Neurosci* 32: 315–46.
- [7] Norman-Haignere S, Kanwisher N, McDermott JH. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J Neurosci* 33: 19451–69.
- [8] Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci USA*. 110:2653–58.
- [9] Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT. (2010). Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci* 30:16643–50.
- [10] Greenlee JDW, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H, Kawasaki H, Howard MA. (2013). Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS One*. 8:e60783.
- [11] Greenlee JDW, Jackson AW, Chen FX, Larson CR, Oya H, Kawasaki H, Chen HM, Howard MA. (2011). Human auditory cortical activation during self-vocalization. *PLoS One* 6: e14744.
- [12] Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82:2346–57.
- [13] Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb Cortex* 15:170–86.
- [14] Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA. (2011). Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. *Cereb Cortex* 21:2332–47.
- [15] Brugge JF, Volkov IO, Oya H, Kawasaki H, Reale RA, Fenoy AJ, Steinschneider M, Howard MA. (2008). Functional localization of auditory cortical fields of human: click-train stimulation. *Hear Res* 238:12–24.
- [16] Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J Neurophysiol* 102:2358–74.
- [17] Howard MA, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724:260–4.
- [18] Nourski KV, Brugge JF, Reale RA, Kovach CK, Oya H, Kawasaki H, Jenison RL, Howard MA. (2013). Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study. *J Neurophysiol* 109:1283–95.
- [19] Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002). Modulation of the auditory cortex during speech: An MEG study. *J Cogn Neurosci* 14:1125–38.
- [20] Romanski LM, Bates JF, Goldman-Rakic PS. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol*. 403:141–57.
- [21] Wolpert DM, Ghahramani Z, Jordan MI. (1995). An Internal Model for Sensorimotor Integration. *Science*. 269:1880–2.
- [22] Brechmann A, Baumgart F, Scheich H. (2002). Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. *J Neurophysiol*. 87:423–33.