SUPPLEMENTARY DATA

Supplementary Material

Methods

The degree to which a measured ERP and its derived spline representation agreed was illustrated (see Fig. 2) in this report using average data obtained at one electrode site (chosen for maximal voltage excursion) within PLST in subject L106 for the first three stimulus types shown in Table 1. Subsequently, the analytical solution to the spline representation of the measured data was used to derive the Spline-Laplacian functional, and the latter used to generate a corpus of Spline-Laplacian waveforms comprising all trials for all stimulus types, analysis windows, and original electrode sites. This corpus of waveforms was then submitted to PCA analysis in order to reduce the dimensionality of the input data vectors; here from 500 to 21. The degree to which a Spline-Laplacian waveform and its PCA-derived representation agree is illustrated in **Figure Appendix-1** using data derived for same electrode site in subject L106 as above. Average Spline-Laplacian waveforms are shown here in blue for unimodal (Auditory or Visual) and bimodal (Audiovisual) speech whereas Average Spline-Laplacian waveforms calculated from the resulting PCA scores and loadings are shown in red. The comparison is representative of the degree to which an analytical solution and its PCA derived reconstruction agree; for most of the waveform the two curves superimpose, with exceptions noted around high-frequency peaks and valleys. Vertical dashed lines mark the temporal boundaries of the three 500 msec analysis windows (AW) included in the Manova analysis. A common ordinate scale is used for all waveforms. Negative polarity in the upward direction.

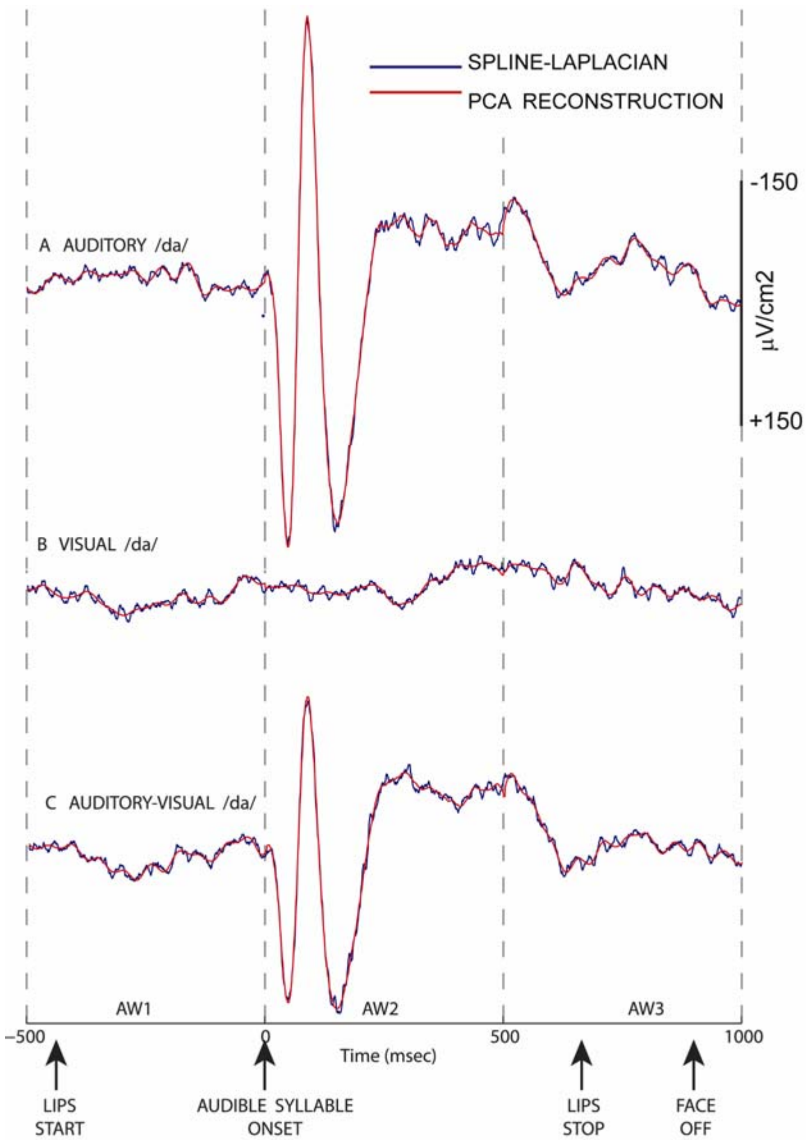

FIGURE APPENDIX-1

Multivariate Statistical Test

Single-factor Manova is a well-known and powerful technique that incorporates the correlations within vector response measurements when detecting an effect among the levels of a single treatment, such as Stimulus Type. Our model, on the other hand, is a multiple-factor Manova (or more properly termed doubly-multivariate repeated measures Manova) as it has two additional factors that represent Electrode Site and Analysis Window. This model enabled us to test hypotheses concerning main effects (e.g. Stimulus Type) and their combinations (e.g. Stimulus Type x Electrode Site x Analysis Window). When there was a strong latter effect, five comparisons (planned contrasts) were employed to identify the specific electrodes and analysis windows contributing to that effect. Performing these multiple comparisons required an adjustment to the raw p-values in order to control for the inflation of the family wise (Type-I) error rate (FWE). The success of this multiplicity correction is suggested by careful examination of the results presented in Figure 5, which shows the proportion of significant electrode sites for each of the contrasts, for each analysis window, and for each of the subjects. Clearly for any given subject, these multiplicity-adjusted counts are at or near zero in AW1 and AW3 even when AW2 contained high counts of significant sites for all five tested contrasts. Without such adjustment, the number of significant sites would have increased dramatically in all three analysis windows.

Results

We interpreted the $\overrightarrow{A_{da}V_{da}}$ vs $\overrightarrow{A_{da}}$ significance maps in this report (e.g. Figs. 6A, 9B&C) as identifying those cortical locations where there was a significant combined influence of vectors $\overrightarrow{V_{da}} + \overrightarrow{INT}$ upon the ECoG. **Figure Appendix-2B** illustrates the average Spline-Laplacian **difference** waveforms for this contrast using the data from subject L106 (see Figures

6B&C). Electrodes sites marked by gray boxes show where the contrast was found significant ($p < 0.05$) since the variation in the difference response was systematically related to the classification factors. **Figure Appendix-2C** shows average Spline-Laplacian waveforms elicited by unimodal Visual speech (V_{da}) for this left-hemisphere subject. As seen here, the $\overrightarrow{V_{da}}$ response field is unremarkable in comparison to these difference waveforms or to response fields obtained with a stimulus that included an audible syllable. This observation suggests that, although we are unable to measure directly the interaction term, the effect we see from this contrast may be carried largely by the interaction term. The Sylvian fissure (SF) and superior temporal sulcus (STS) are shown in yellow on the spatial maps of waveforms. The ordinate scale (vertical line: -240 to +340 mV/cm2) is common to both maps and the abscissa scale includes all three analysis windows.

Figure Appendix-2E illustrates the average Spline-Laplacian **difference** waveforms for this contrast using the data from subject R104 (see Figures 5&8). The single electrodes site marked by a gray box shows where the contrast was found significant. The paucity of such sites was typical of right-hemisphere cases. **Figure Appendix-2F** shows average Spline-Laplacian waveforms elicited by unimodal Visual speech (V_{da}) for this right-hemisphere subject. Again, the $\overrightarrow{V_{da}}$ response field is generally unremarkable in comparison to these difference waveforms or to response fields obtained with a stimulus that included an audible syllable.

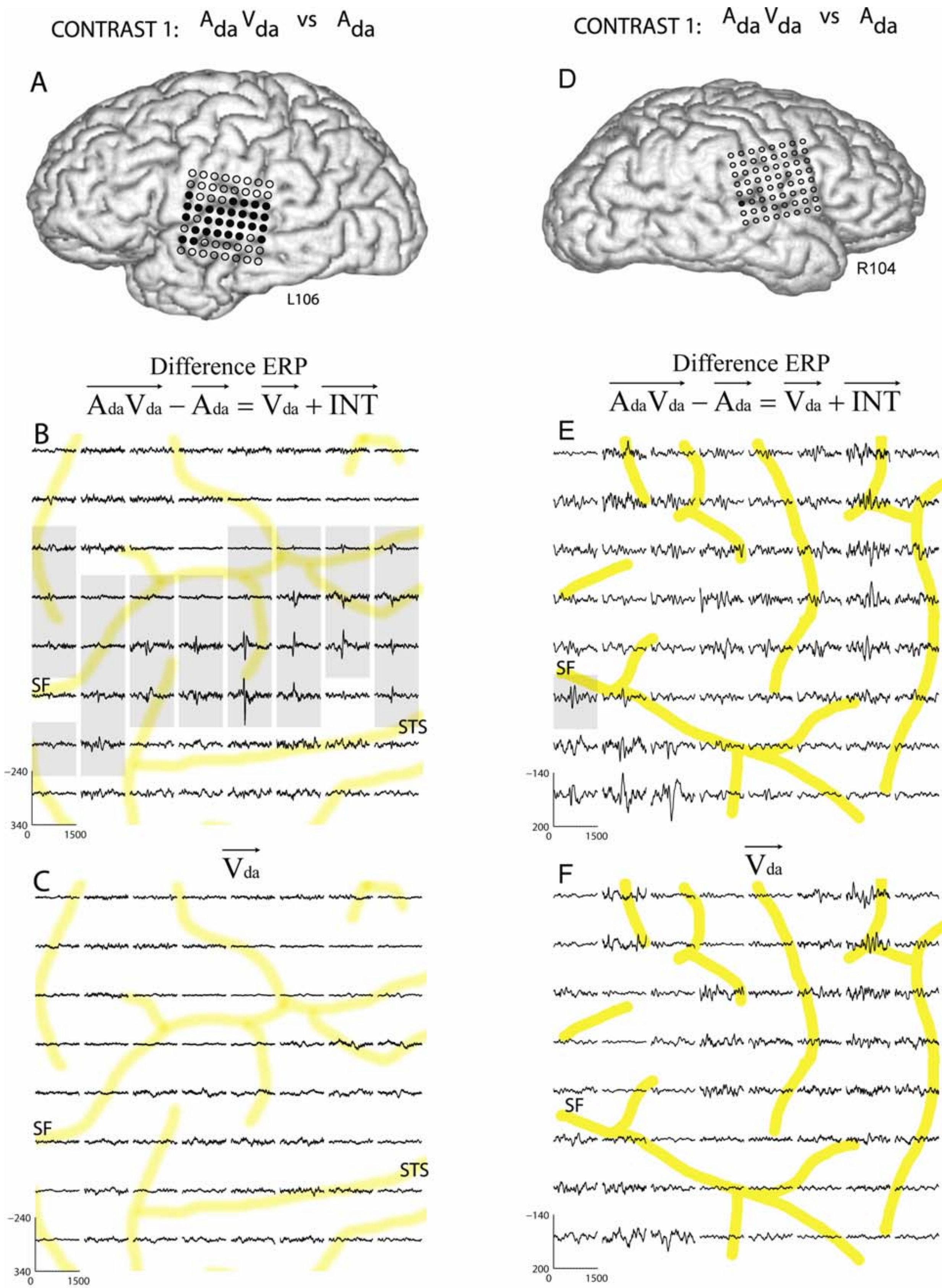

FIGURE APPENDIX-2

We interpreted the $\overrightarrow{A_{da}V_{da}}$ vs $\overrightarrow{V_{da}}$ significance maps in this report (e.g. Figs. 7A, 8D-F) as representing those cortical locations where the combined influence of vectors $\overrightarrow{A_{da}} + \overrightarrow{INT}$ upon the ECoG was detected reliably. **Figure Appendix-3B** illustrates the average Spline-Laplacian **difference** waveforms for this contrast using the data from subject L106 (see Figures 7B&C). Electrodes sites marked by gray boxes show where the contrast was found significant ($p < 0.05$) since the variation in the difference response was systematically related to the classification factors. These difference waveforms by themselves, however, do not permit parsing the significance effect between the two terms $\overrightarrow{A_{da}}$ and $\overrightarrow{INT}$. **Figure Appendix-3C** shows average Spline-Laplacian waveforms elicited by unimodal Auditory speech (A_{da}) for this left-hemisphere subject. As seen here, auditory alone (A_{da}) stimulation produced a clearly distinguishable response field in this subject and all others in this study. This observation suggests that, although we are unable to measure directly the interaction term, the effect we see from this contrast may be interpreted as arising mainly from the $\overrightarrow{A_{da}}$ term.

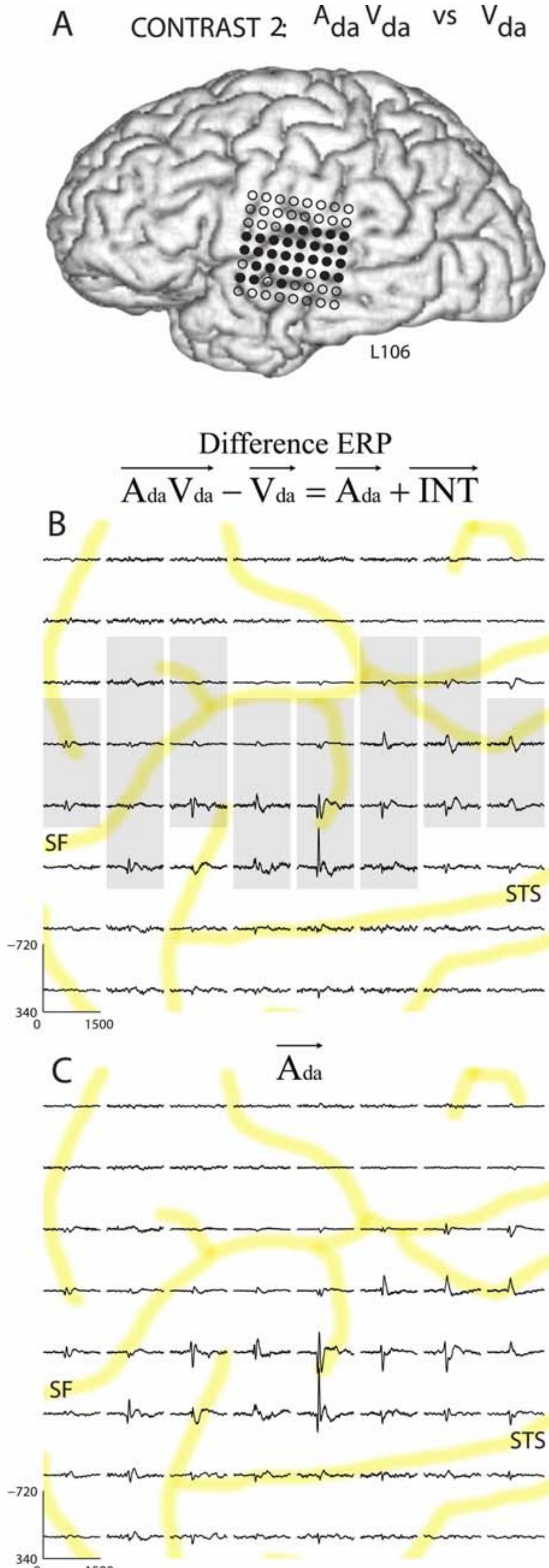

FIGURE APPENDIX-3

The Laplacian transformation of Spline-fit ERPs is not expected to be very accurate near the edges of our electrode grid since this spatial sample is limited to the lateral temporal cortex. We attempted to reduce this edge effect during spline interpolation (not entirely successfully) by adding extra electrode sites along each edge of the recording grid (not shown) and requiring these extra waveform voltages to be zero at each sample time. Here we illustrate in **Figure Appendix-4A** average ERPs at each of the original recording sites derived from the spline fit with zero padding and in **4B** without such padding. Ordinate scale –100 to 100 mV. The difference between the two maps (**4C**) is small even near the edges even near the edges of the recording grid; ordinate scale –10 to 10 mV. A similar pattern can be seen when these Spline-ERP maps are replaced by their Spline-Laplacian counterparts in the **Figure Appendix-4D,E,F**. The Laplacian transformation involves a second derivative of the Spline-fit, the inaccuracy near the edges of the electrode grid (**4F)** is exaggerated relative to their spline counterparts. Comparing the statistical results obtained when zero-padding was used and when it was not used, did not prove consequential to the conclusions obtained in this study.

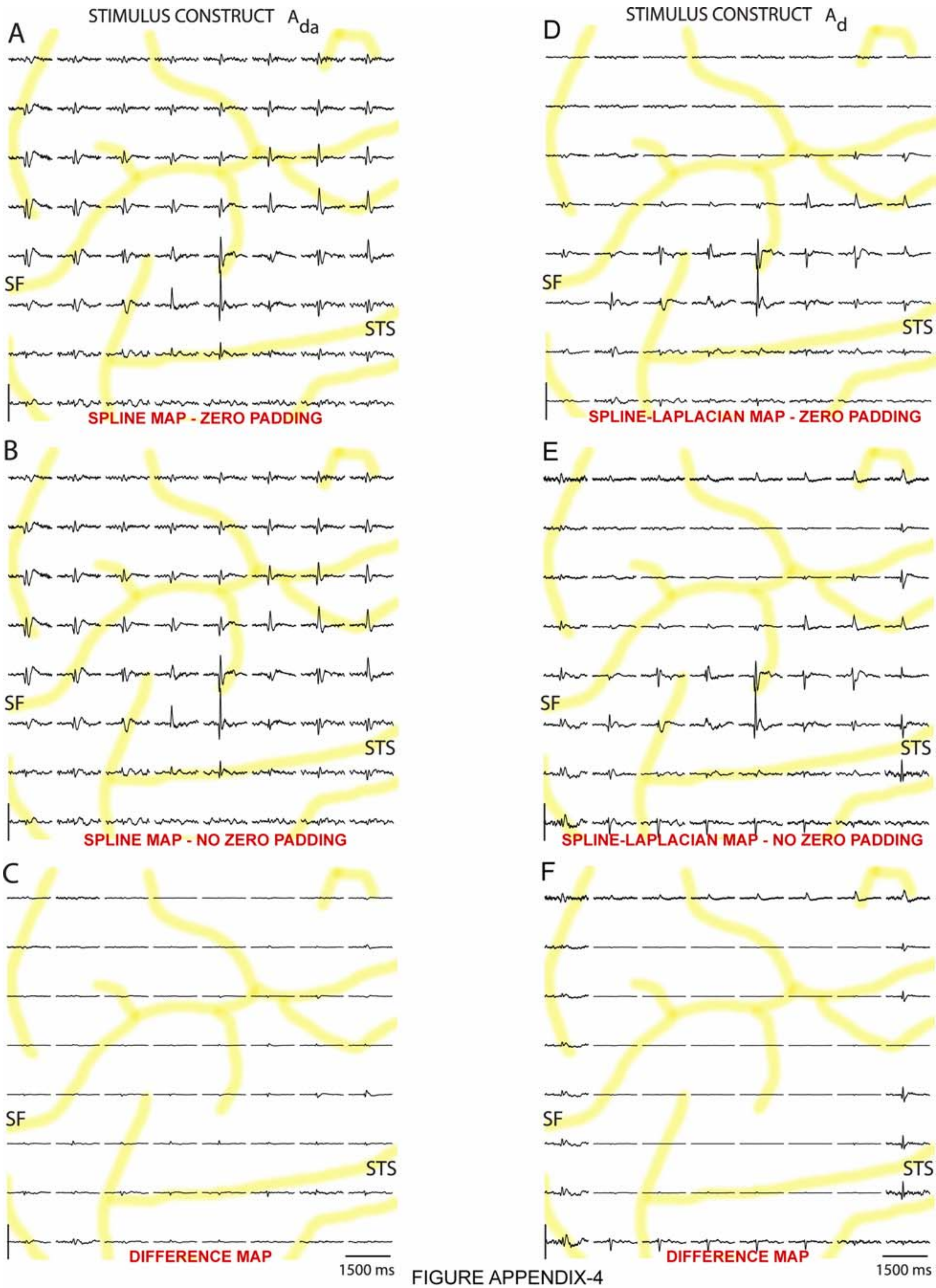

FIGURE APPENDIX-4

Discussion

Our data analyses differed from those carried out in many non-invasive electrophysiological and functional neuroimaging studies of AV speech activation of human cortex. In those studies, AV interactions were generally inferred from a comparison between a bimodal ***univariate*** response and one (the largest) or more (sum or average) of the ***univariate*** responses to the unimodal components of an AV stimulus (Beauchamp et al., 2004, 2005; Calvert et al., 2000; Giard and Peronnet, 1999; Meredith and Stein, 1986; Van Wassenhove et al., 2005; Wright et al., 2003). The attempt to measure directly an AV interaction ($\overrightarrow{INT}$) using this approach involves simply calculating the difference between the bimodal (*pair*) response ($\overrightarrow{AV}$) and the *sum* of the unimodal responses ($\vec{A} + \vec{V}$). The outcome of this calculation has commonly shown that AV interactions may be 'additive' or 'subtractive' in nature depending upon choice and size of the analysis window and perhaps upon other exogenous variables, such as the listener's attention and task performance (Beauchamp et al., 2004, 2005; Calvert and Thesen, 2004; Fort and Giard, 2004; Ghazanfar et al., 2005; Laurienti et al., 2005; Olson et al., 2002; Teder-Salejarvi et al., 2002; Wright et al., 2003). In our data analysis, the dependent ERP measurement was ***multivariate*** and our statistical test did not depend upon the measurement of a single deflection in the ERP nor did it depend upon a simple difference in the response variable. Rather, any systematic difference between the contrasted ERPs, beyond chance, is sufficient to mark an effect (see Online Supplementary Data: Results). Thus the terms 'additive' or 'subtractive' are not appropriate in the present study. Importantly, we've made no specific hypothesis concerning the additive or subtractive nature of a possible effect given the operational dependencies that such adjectives imply. In this sense our position regarding ERP analyses is essentially that taken recently by Beauchamp (2005) concerning complementary fMRI data, namely there is no general agreement on the optimal criteria to be used to classify brain regions as ‘multisensory’.

The ‘sum’ vs. ‘pair’ approach, while attractive in its simplicity, has a serious and well-know potential shortcomings (Beauchamp et al., 2004, 2005; Calvert and Thesen, 2004; Fort et al., 2002; Fort and Giard, 2004; Teder-Salejarvi et al., 2002). In particular, if there is a unknown (e.g. to attention, arousal etc) potential ($\vec{U}$) that is common to all three stimulus types, then the difference calculation of [{ $\overrightarrow{AV} + \vec{U}$ - { $\vec{A} + \vec{U} + \vec{V} + \vec{U}$ }], yields two terms ($\overrightarrow{INT} + \vec{U}$), instead of a single interaction term. A number of analysis modifications have been suggested or implemented to ameliorate this shortcoming, including making comparisons only within an early analysis window that is reasoned to be free of the unwanted and unknown potential (Calvert and Thesen, 2004; Fort et al., 2002; Foxe et al., 2000; Giard and Peronnet, 1999; Molholm et al., 2002).

Our experimental design and analytical approach circumvented the latter problem because the unknown potential cancels out in all of the difference calculations shown in Table 2. In addition, rather than represent each ERP by a scalar measurement, we have represented an ERP by a vector. In our calculations, the interaction response vector $\overrightarrow{INT}$ in the remainder is always accompanied by a unimodal response vector ($\vec{A}$ or $\vec{V}$). Thus, we chose this approach because it allowed us to take into account the relative contributions of two vectors in the interpretation of AV responses without having to consider some unknown potential contributing to the ERP, and because we had no *a priori* evidence that restricted the dependent measure to a single time point (or narrow time window) in the evoked waveform.

References

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809-823.

Beauchamp MS (2005) Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3:93-113.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649-657.

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris 98:191-205.

Fort A, Delpuech C, Pernier J, Giard MH (2002) Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. Cereb Cortex 12:1031-1039.

Fort A, Giard M-H (2004) Multiple electrophysiological mechanisms of audiovisual integration in human perception. In: The Handbook of Multisensory Processes(Calvert, G. A. et al., eds), pp 503-514 Cambridge: MIT Press.

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res 10:77-83.

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278-285.

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473-490.

Laurienti P, Perrault T, Stanford T, Wallace M, Stein B (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. Exp Brain Res 166:289-297.

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640-662.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res 14:115-128.

Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res Cogn Brain Res 14:129-138.

Teder-Salejarvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106-114.

van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci U S A 102:1181-1186.

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13:1034-1043.