

# Human Auditory Cortex Activity Shows Additive Effects of Spectral and Spatial Cues during Speech Segregation

In noisy social gatherings, listeners perceptually integrate sounds originating from one person's voice (e.g., fundamental frequency ( $f_0$ ) and harmonics) at a particular location and segregate these from concurrent sounds of other talkers. Though increasing the spectral or the spatial distance between talkers promotes speech segregation, synergetic effects of spatial and spectral distances are less well understood. We studied how spectral and/or spatial distances between 2 simultaneously presented steady-state vowels contribute to perception and activation in auditory cortex using magnetoencephalography. Participants were more accurate in identifying both vowels when they differed in  $f_0$  and location than when they differed in a single cue only or when they shared the same  $f_0$  and location. The combined effect of  $f_0$  and location differences closely matched the sum of single effects. The improvement in concurrent vowel identification coincided with an object-related negativity that peaked at about 140 ms after vowel onset. The combined effect of  $f_0$  and location closely matched the sum of the single effects even though vowels with different  $f_0$ , location, or both generated different time courses of neuromagnetic activity. We propose that during auditory scene analysis, acoustic differences among the various sources are combined linearly to increase the perceptual distance between the co-occurring sound objects.

**Keywords:** attention, MEG, scene analysis, streaming, speech

## Introduction

*f* *f*  
*f* *f*  
*f*

%

*f*

*f* *f*

~

~

*f*

*f*

*f*

$\Delta f$

*f*

$\Delta f$   
*f*

*f*

*f*

*f*

%

*f*

## Material and Methods

### Participants

= ±

”

3

” ” ”

= ±

### Data Acquisition

### Stimuli and Task

3

*f*

±

### Data Analysis

*f*

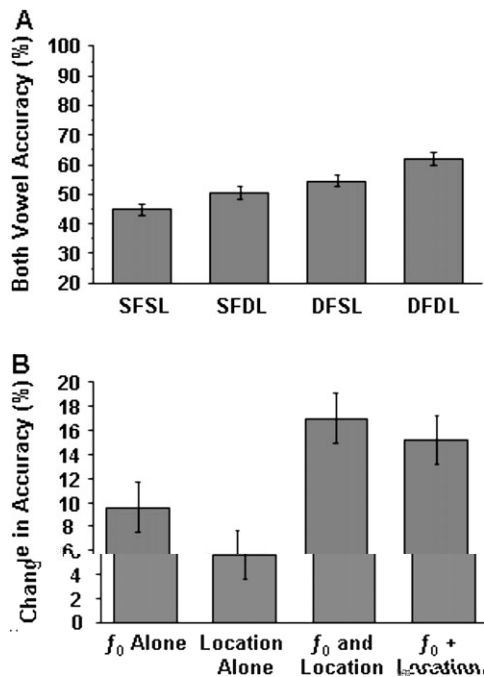
$f$

$f$

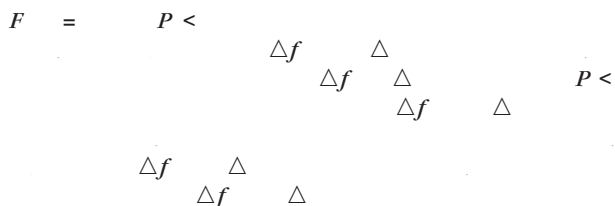
***Dipole Source Analysis***

%

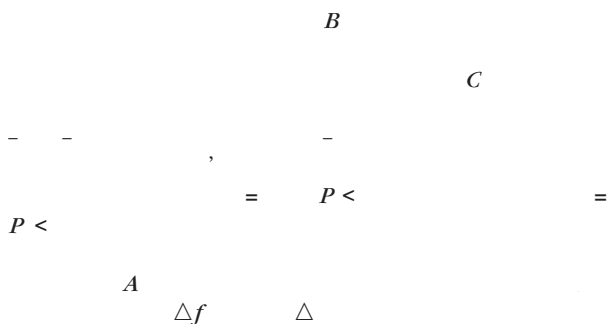
=



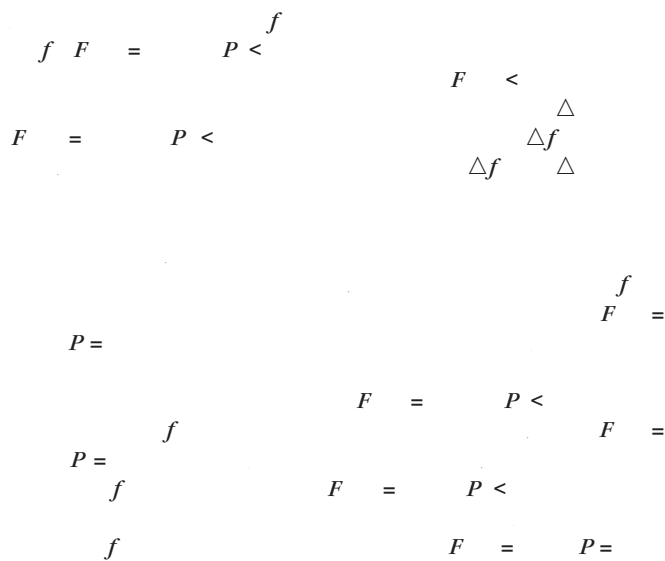
**Figure 2.** Behavioral performance and behavioral benefit in Experiment 2. (A) Proportion of trials in which both vowels were correctly identified under 4 stimulus types: SFSL, SFDL, DFSL, and DFDL. (B) Changes in accuracy of identification of both vowels (compared with performance when 2 vowels shared same  $f_0$  and location) for stimulus conditions with only  $\Delta f_0$  ( $f_0$  alone),  $\Delta$ location (location alone), and both  $\Delta f_0$  and  $\Delta$ location ( $f_0$  and location). A linear sum of change by  $\Delta f_0$  alone and that by  $\Delta$ location alone is shown also ( $f_0 +$  location). The error bars ( $\pm$ standard error of the mean) indicate the within-subject variability for each condition.



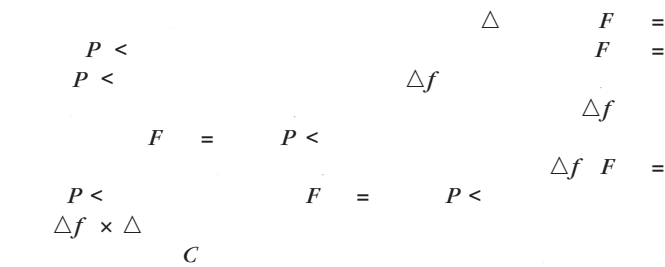
*Distance Wa ef . m*  
A



A



*Objec -Rela ed Nega i i*



C

$$P < \Delta f \quad \Delta \quad \dots = - \quad P < \\ \Delta f \quad \Delta \quad \dots = - \quad P >$$

C

$$\Delta \quad - \quad P = \quad \Delta f$$

$$F = \quad P =$$

$$\Delta f \quad - \quad \Delta \quad - \quad \Delta f \quad \Delta$$

$$F <$$

ER-SAM S . ce Ac i i

f

f

f

$$P < \\ P <$$

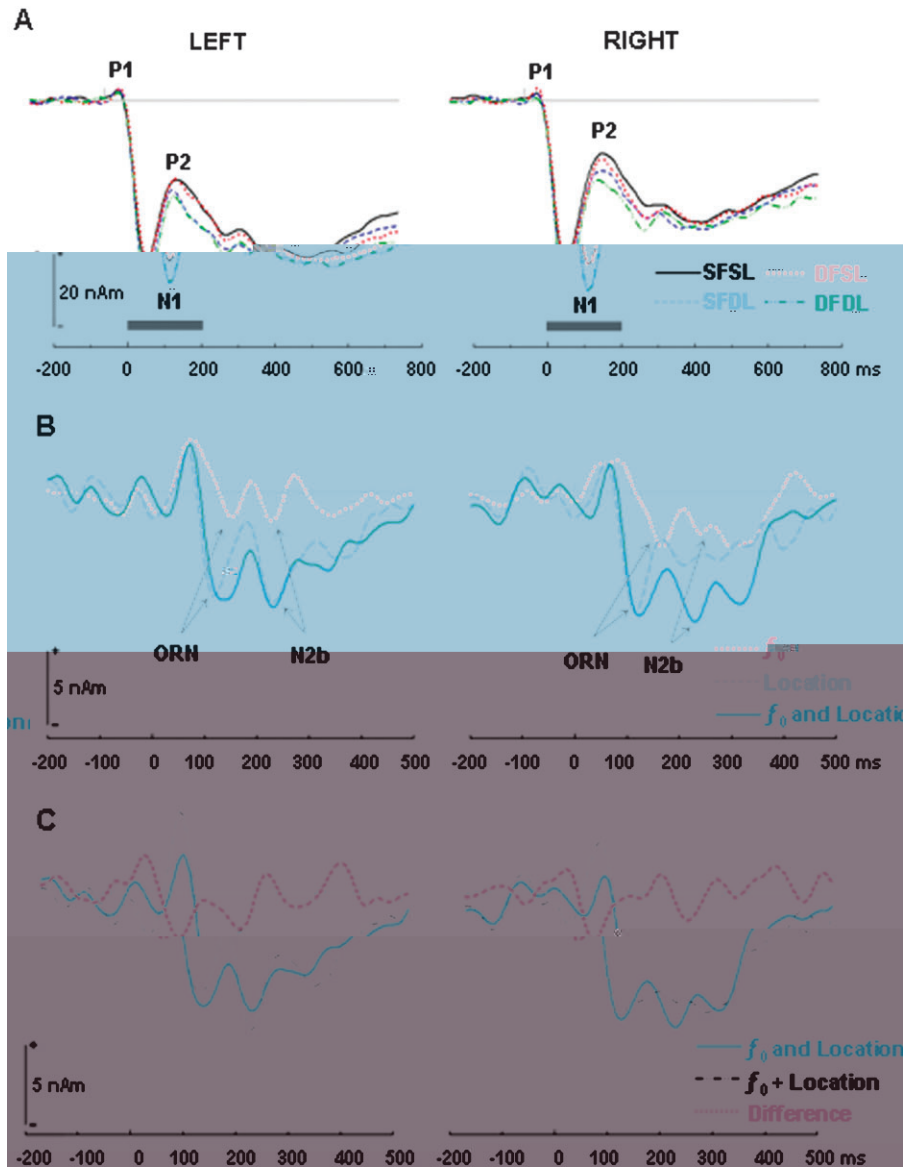
$$\Delta f \quad \Delta$$

B ain-Beba i . C . . ela i n

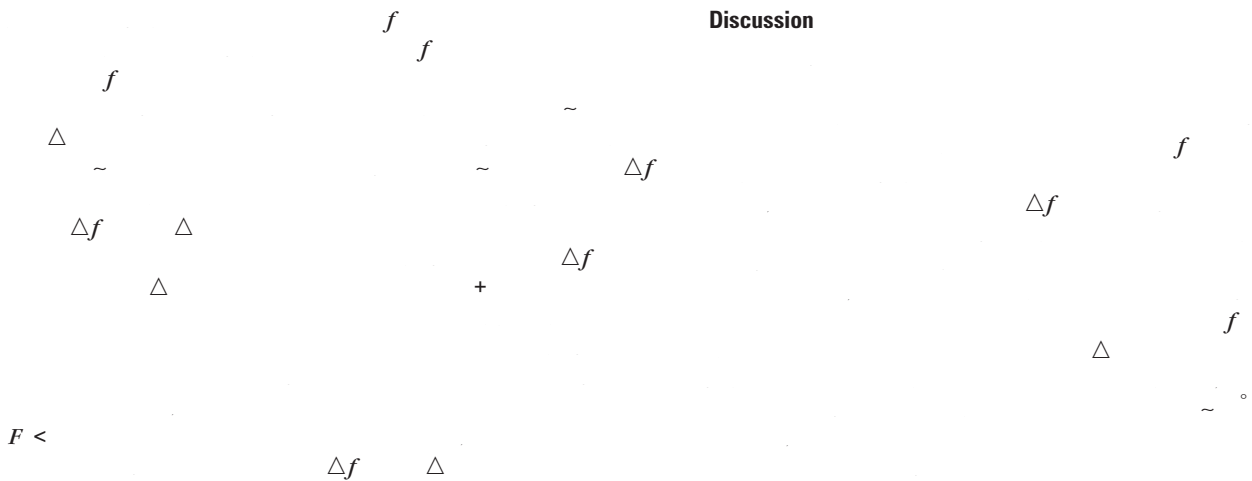
$$F = \quad P = \quad \Delta f \quad \Delta$$

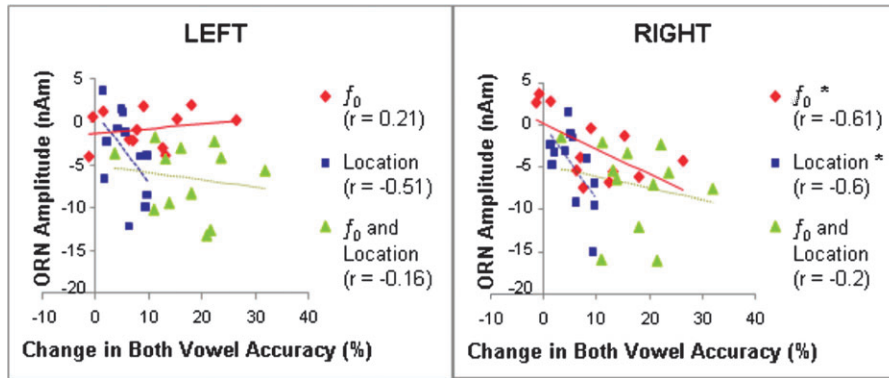
$$\Delta f \quad \Delta$$

$$\Delta f \quad \dots = -$$

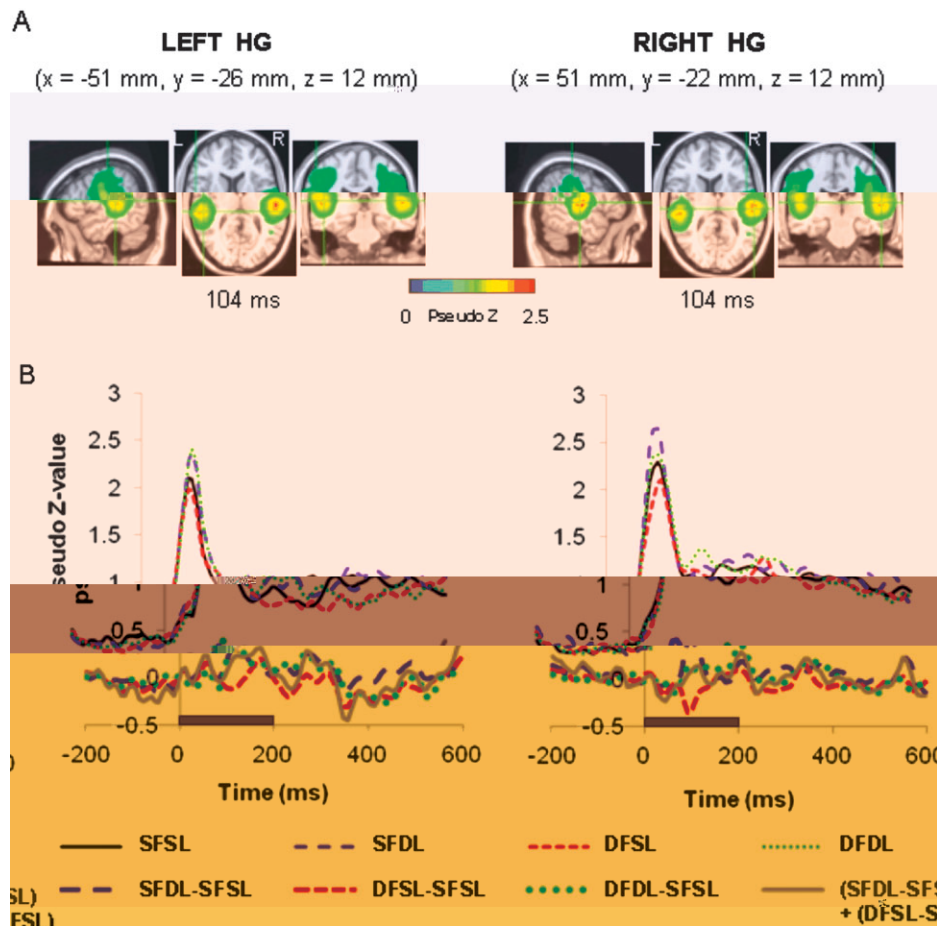


**Figure 4.** Group mean dipole source waveforms. (A) Group mean source waveforms for AEFs for 4 stimulus conditions. The gray rectangles represent the duration of the double-vowel stimulus. (B) The differences in source waveforms between stimuli with same  $f_0$  and same location and stimuli with  $\Delta f_0$  and/or  $\Delta$ location. (C) Comparison between difference waveforms for both  $\Delta f_0$  and  $\Delta$ location ( $f_0$  and location) and a linear sum of difference waveforms for  $\Delta f_0$  alone and that for  $\Delta$ location alone ( $f_0$  + location).

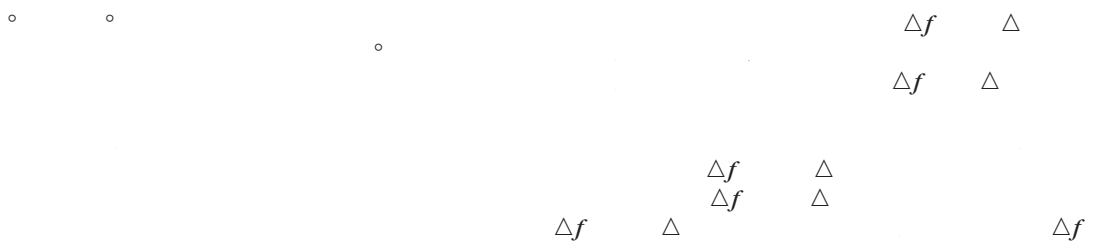




**Figure 5.** Brain-behavior correlation. Individual changes in source waveform amplitude during the ORN interval (140–180 ms) are plotted against the listeners' changes in accuracy of identification of both vowels for stimuli with  $\Delta f_0$  alone ( $f_0$ ),  $\Delta$ location alone (location), and both  $\Delta f_0$  and  $\Delta$ location ( $f_0$  and location). \* $P < 0.05$ .



**Figure 6.** ER-SAM maps and source waveforms in left and right Heschl's gyri. (A) Thresholded group-mean ER-SAM maps for SFSL condition at N1m latency of 104 ms. (B) Time courses of source activities associated with 4 stimulus conditions (SFSL, SFDL, DFSL, and DFDL) and 4 differences between conditions [(SFDL-SFSL), (DFSL-SFSL), (DFDL-SFSL), (SFDL-SFSL) + (DFSL-SFSL)]. Pseudo Z values represent the ratio of signal-to-noise power of the evoked response. Note that in both hemispheres, the difference in source waveforms for both  $\Delta f_0$  and  $\Delta$ location (DFDL-SFSL) closely matches the linear sum of  $\Delta f_0$  and  $\Delta$ location alone [(SFDL-SFSL) + (DFSL-SFSL)].



$\Delta$

”

$\Delta f$   $\Delta f$   $\Delta$   
 $\Delta$

$\Delta f$   $\Delta$

$\Delta f$

$\Delta f$   $\Delta$

$f$

$\Delta$

$\Delta f$

$\Delta$   $\Delta f$

$\Delta$

$\Delta f$   $\Delta$

$f$



**Funding**

”

+

+ + <

**Notes**

*C n i c f I n e e*

**References**

”

”

