Supplementary Motor Area Activation in Disfluency Perception. An fMRI Study of Listener Neural Responses to Spontaneously Produced Unfilled and Filled Pauses

Robert Eklund[1] & Martin Ingvar[2, 3]

[1] Department of Culture and Communication, Linköping University, Linköping, Sweden
[2] Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
[3] Osher Center for Integrative Medicine, Karolinska Institute, Stockholm, Sweden

ABSTRACT

Spontaneously produced Unfilled Pauses (UPs) and Filled Pauses (FPs) were played to subjects in an fMRI experiment.

For both stimuli increased activity was observed in the Primary Auditory Cortex (PAC). However, FPs, but not UPs, elicited modulation in the Supplementary Motor Area (SMA), Brodmann Area 6.

Our results provide neurocognitive confirmation of the alleged difference between FPs and other kinds of speech disfluency and could also provide a partial explanation for the previously reported beneficial effect of FPs on reaction times in speech perception.

Our results also have potential implications for two of the suggested functions of FPs: the "floor-holding" and the "help-me-out" hypotheses.

Keywords: speech disfluency, filled pauses, unfilled pauses, speech perception, spontaneous speech, fMRI, Auditory Cortex, PAC, Supplementary Motor Area, SMA, Brodmann Area 6, BA6

1

INTRODUCTION

Almost no one is completely fluent when speaking.

Speech disfluency has been studied since at least the 1930s. Disfluencies have commonly been regarded as "performance errors" in speech production.

However, several studies indicate that certain kinds of disfluencies can have beneficial effects on listener perception (Fraundorf & Watson, 2011; Barr & Seyfeddinipur, 2010; Ferreira & Bailey, 2004; Fox Tree, 2001, 1995; Reich, 1980).

The most common voiced disfluency is the filled pause (FP), "eh". The reported average frequency of filled pauses (FPs) ranges from 1.9 to 7.6% at word level (Eklund, 2010).

Neurocognitive studies on disfluency have traditionally consisted of electrophysiological studies, starting with Kutas and Hillyard (1980) on the N400 component.

The present study uses functional Magnetic Resonance Imaging (fMRI) to analyse the effect of authentic disfluencies proper to study the effect of unfilled and filler pauses on brain processing.

2

METHOD (1)

Subjects

The subjects were 16 healthy adults (9 F/7M) ages 22–54 (mean age 40.3, standard deviation 9.5) with no reported hearing problems. All subjects were right-handed as determined by the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects possessed higher education. Following a description of the study, including a description of fMRI methodology, written and informed consent was obtained from all subjects. A small participation remuneration was also administered.

Equipment

The fMRI scanner used was the General Electric 1.5T Excite HD Twinspeed scanner at Karolinska Institute, MR-center, Stockholm, Sweden. The coil used was a General Electrics Standard bird-cage head-coil (1.5T).

Stimulus data

The stimulus data used were excerpts from the human–human dialog speech data described in Eklund (2004: 187–190). Subjects were asked to play the role of travel agents listening to customers making travel bookings over the telephone, following a task sheet (Eklund, 2004: 185)

From the original data set, four speakers were chosen (2M/2F) and a number of sentences were excised that were fluent except that they included a number of UPs and an approximately equal number of FPs. The resulting number of both UPs and FPs roughly corresponded to reported incidence of UPs and FPs in spontaneous speech.

Stimulus data are shown in Table 1 below.

Table 1. Stimulus data. Legend: UPs = Unfilled Pauses; FP = Filled Pauses; MIT = Mean Interstimulus Time is given in seconds.

Stimulus File	No. UPs / MIT	No. FPs / MIT
1	17 / 11.9 s	23 / 7.1 s
2	9 / 9.7 s	8 / 13.8 s
3	10 / 5.5 s	9 / 8.7 s
4	22 / 7.2 s	15 / 10.7 s

3

METHOD (2)

Experimental design

The stimulus files described above were used in an event-related experiment. After initial localizer anatomical scanning sessions, the four stimulus files (M/F/M/F) were played. During the intermissions the subjects were briefed whether they were still awake/focused on the task. Interstimulus intervals were of sufficient duration so as to allow for reliable BOLD acquisition. FPs and UPs were modelled as events in SPM and were convolved using the Haemodynamic Response Function (HRF) in SPM.

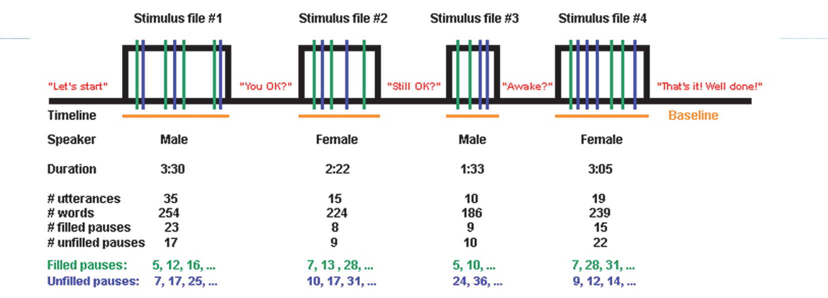
Experimental setting

The subjects lay supine/head first in the scanner with earplugs to protect them from scanner noise and headphones with the sound data played to them. The perceived sound level was quite sufficient and no subjects reported having any problems hearing what was said. Head movement was constrained using foam wedges and/or tape.

Experimental instructions

The subjects were instructed to listen carefully to what was said, as if they were the addressed travel agent, but that they were not expected to react verbally to the utterances or say anything, only that they needed to pay attention to the information provided by the clients. All subjects understood the instructions without any confusion.

Experimental design: data acquisition



4

METHOD (3)

Post-experiment interview

After the scanning, all subjects were interviewed in order to confirm that they had been awake and focused during the experiment. A self-rating scale of how attentive the subjects felt they had been during the sessions was also used. All subjects reported that they had been attentive at a satisfactory level.

MRI scans

For each subject, a T1-weighted coronal spoiled gradient recalled (SPGR) image volume was obtained to serve as anatomical reference (TR/TE= 24.0/ 6.0 ms; flip angle 35°; voxel size = $1 \times 1 \times 1$ mm^3).

Moreover, BOLD-sensitized fMRI was carried out by using echo-planar imaging EPI using 32 axial slices (TR/TE=2500/40 ms, flip = 90 deg, voxel size = $3.75 \times 3.75 \times 4$ mm^3).

In total, T2*-weighted images were collected four sessions: (3m30s/80 volumes; 2m22s/53 volumes; 1m33s/33 volumes; 3m05s/70 volumes).

Post-processing

The images were post-processed and analyzed using MatLab R2007a and SPM5 (Friston et al., 2007). The images were realigned, co-registered and normalized to the EPI template image in SPM5 and finally smoothed using a FWHM (Full-Width Half Maximum) of 6 mm.

Regressors pertaining to subject head movement (3 translational and 3 rotational) were included as parameters of no-interest in the general linear model at the first level of analysis.

No subjects were excluded due to head motion or for any other image acquisition related causes. Analyses were also carried out using the SPM Anatomy Toolbox (Amunts, Schleicher & Zilles, 2007; Eickhoff et al., 2007, 2006, 2005).

5

COMMENTS ON DATA AND ANALYSES

1. We used *fMRI to study disfluency perception*, not EEG with its focus on temporal aspects of speech perception.
2. We investigated perceptual modulation caused by *FPs proper*, not their effect on ensuing items (words, phrases) or general cognitive processing.
3. Unlike previous studies where the auditory stimuli often have been scripted laboratory speech, we used *ecologically valid stimulus data.*

ANALYSES AND RESULTS (1)

Using Fluent Speech (FS) as the baseline condition, the following three contrasts were analyzed:

(1) Filled Pauses > Fluent Speech

(2) Unfilled Pauses > Fluent Speech

(3) Filled Pauses > Unfilled Pauses

The results were calculated with a False Discovery Rate at $p < 0.05$ (Genovese, Lazar & Nichols, 2002) with a cluster level threshold of 10 contiguous voxels.

No activation in Brodmann Area 22, associated with semantic processing, was observed.

6

ANALYSES AND RESULTS (2)

Results are shown in Table 2 and Figures 1–3 below:

Table 2. Locations of significant activation for three contrasts, FDR-corrected at $p < 0.05$ and with a cluster level threshold of 10 contiguous voxels, analyzed with SPM (Friston et al., 2007) and the SPM Anatomy Toolbox (Eickhoff et al., 2007, 2006, 2007). Brodmann Areas were identified using the Talairach Atlas (Talairach & Tournoux, 1988) and the Talairach Client/Daemon (**www.talairach.org**), using a Cube Range setting of ±5 mm.

Contrast	Area	Brodmann	Coordinates			Number of voxels	T
			x	y	z		
Filled Pauses > Fluent Speech	Auditory cortex (left)	BA 40/41	−42	−27	+11	5881	11.08
	Auditory cortex (right)	BA 41/42	+53	−18	+6	4591	14.75
	Cerebellum (left)	–	−24	−71	−26	932	7.77
	Putamen (left)	–	−22	+4	+1	684	9.13
	Supplementary Motor Area (left)	BA 6	−47	−8	+55	285	6.64
	Inferior Frontal Gyrus (right)	BA6	+46	+5	+28	147	5.36
	Supplementary Motor Area (right)	BA 6	+50	−4	+54	103	5.31
	Supplementary Motor Area (medial)	BA 6	+8	+14	+49	101	6.17
	Supplementary Motor Area (medial)	BA 6	+3	+2	+61	63	3.74
Unfilled Pauses > Fluent Speech	Auditory cortex (right)	BA 41/42	+58	−22	+7	1499	10.81
	Auditory cortex (left)	BA 41/42	−57	−26	+10	756	8.54
	Rolandic Operculum (right)	BA 44	+57	+1	+22	31	6.48
Filled Pauses > Unfilled Pauses	Auditory cortex (left)	BA 41/42	−54	−27	+9	1517	9.95
	Auditory cortex (right)	BA 21/22/41	+60	−16	+2	884	10.03
	Cerebellum (left)	–	−21	−67	−33	423	6.51
	Supplementary Motor Area (right)	BA 6	+7	+13	+49	60	4.86
	Inferior Frontal Gyrus (left)	BA 44/45/47	−46	+13	+2	25	5.06
	Supplementary Motor Area (left)	BA 6	−7	+2	+55	16	5.17

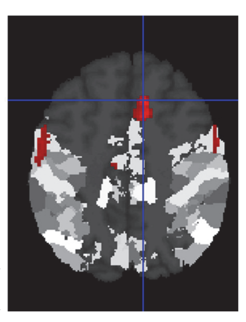


Figure 1. FPs > FS. 101 voxels. 10% of cluster in right area 6.

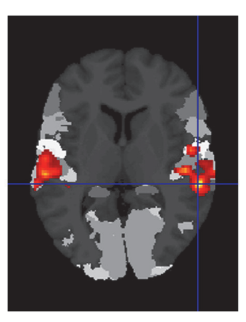


Figure 2. UPs > FS. 1499 voxels. 10.3% of cluster in right TE 3. 9.2% of cluster in right TE 1.0. 5.9% of cluster in right OP 1. 5.6% of cluster in right TE 1.2.

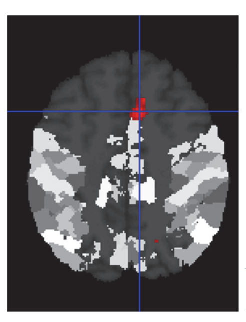


Figure 3. FPs > UPs. 60 voxels. 13.1% of cluster in right area 6.

7

DISCUSSION AND CONCLUSIONS

(1) **UPs and FPs modulated the Primary Auditory Cortex** (PAC). That heightened attention influences auditory perception has been shown (Petkov et al. 2004; and more) and the observed attention-heightening function of FPs could possibly help explain the shorter reaction times to linguistic stimuli that follow FPs as reported by e.g. Fox Tree (2001, 1995). Note, however, that UPs also exhibited PAC activation.

→ Both UPs and FPs lead to heightened attention.

(2) **Filled Pauses activated (pre-)motor areas**, (e.g. BA6). (Not observed in UPs.)

Known already since Brickner (1940) that SMA is active in the processing of speech, and several later studies have confirmed both SMA and pre-SMA play a role in both speech *production* (e.g. Goldberg, 1985; Alario et al., 2006) and speech *perception* (e.g. Iacoboni, 2008; Wilson et al., 2004 and many more).

→ FPs seem to "kick-start" speech production.

8

POTENTIAL IMPLICATIONS

(1) **"floor-holding" hypothesis** of FPs, as first proposed by Maclay and Osgood (1959): that FPs are used to keep the floor in polylogs.

→ Our results imply that FPs could possibly be counter-productive for this.

(2) **"help-me-out" hypothesis**, as suggested in Clark and Wilkes-Gibbs (1986): FPs are/can be used as a (semi-deliberate) signal asking for interlocutor help in conversation

→ Our results imply that FPs might serve this function well.

SUMMING IT UP

Our results – confessedly speculative – suggest that FPs, but not FS/UPs, activate motor areas in the listener brain.

Both FPs/UPs activate PAC, which lends support to the attention-heightening hypothesis that has been forwarded in the literature and it would also seem clear that it is not the break in the speech stream per se that causes this activation, since UPs seemingly do not have this effect.

9