Neural correlates of the processing of unfilled and filled pauses

Robert Eklund[1], Peter Fransson[2], 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
robert.eklund@liu.se, peter.fransson@ki.se, martin.ingvar@ki.se

ABSTRACT

Spontaneously produced Unfilled Pauses (UPs) and Filled Pauses (FPs) were played to subjects in an fMRI experiment. While both stimuli resulted in increased activity in the Primary Auditory Cortex, FPs, unlike UPs, also elicited modulation in the Supplementary Motor Area, Brodmann Area 6. This observation provides neurocognitive confirmation of the oft-reported difference between FPs and other kinds of speech disfluency and also could provide a partial explanation for the previously reported beneficial effect of FPs on reaction times in speech perception. The results are discussed in the light of the suggested role of FPs as floor-holding devices in human polylogs.

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

A characteristic of spontaneous spoken language is that almost no one is completely fluent, and the most common voiced disfluency is the filled pause, "eh". The reported average frequency of filled pauses (FPs) ranges from 1.9 to 7.6% (Eklund, 2010).

While a commonly expressed view regards speech disfluency as errors in speech production, there are several studies that 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).

While several behavioral studies of speech disfluency have been carried out over the years, only a few strictly neurocognitive studies have been performed, mostly using electrophysiological methods and/or scripted or enacted disfluencies. Moreover, most of these studies have focused on the effect of speech disfluency on higher-level speech processing, like syntactic parsing. 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

2.1. 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. After 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.

2.2. 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).

2.3. 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.

Table 1. Stimulus data. Legend: UPs = Unfilled Pauses; FPs = Filled Pauses; MIT = Mean Interstimulus Time 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
3	10 / 5.5 s	9 / 8.7 s
4	22 / 7.2 s	15 / 10.7 s

2.4. 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.

2.5. 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.

2.6. 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.

2.7. 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.

2.8. 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).

2.9. 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).

3. ANALYSES AND RESULTS

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 (FDR) at $p < 0.05$ (Genovese, Lazar & Nichols, 2002) with a cluster level threshold of 10 contiguous voxels.

First, no activation in BA22, associated with semantic processing, was observed.

For **FP > FS**, increased activation was found in Primary Auditory Cortex (PAC) (Morosan et al., 2001; Rademacher et al., 2001) bilaterally, and in subcortical areas (cerebellum, putamen) and most interestingly in the Supplementary Motor Area (SMA), Brodmann Area 6 (BA6). Activation was also observed in the Inferior Frontal Gyrus (IFG). Typical BA6 modulation is shown in Figure 1.

For **UP > FS**, increased activation was observed in PAC, bilaterally, and in Heschl's Gyrus, the Rolandic Operculum and BA44. We did not observe any activation of SMA.

For **FP > UP**, the activation was very similar to that of FPs over FS. This suggests that FPs and UPs equally affect attention in the listener, but while UPs modulates syntax processing areas, this is not the case for FPs that instead modulate motor areas in the perceiving brain. Also, from the point of view of FPs, fluent speech and UPs seem to constitute more or less the same phenomenon in that there is no observed difference between the two contrasts FPs > FS and FPs > UPs.

4. DISCUSSION

As has already been pointed out, we also focused our analyses on FPs/UPs proper instead of their effect on subsequent linguistic items (e.g. words). The effect on subsequent words following FPs has been studied with electrophysiology, e.g. N400 attenuation on the word following an FPs, as reported in Corley, McGregor and Donaldson (2007). Our main observations will be presented in the following sections.
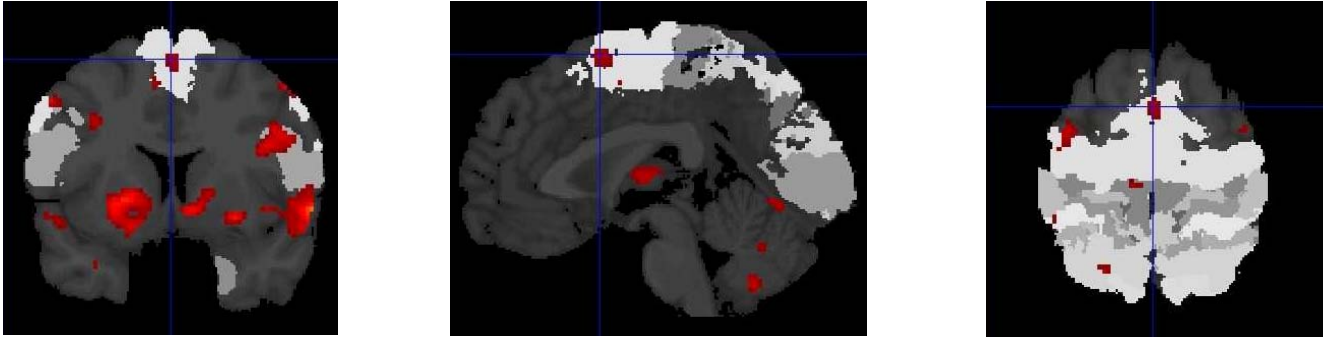
Figure 1. Observed modulation for the contrast Filled Pauses greater than Fluent Speech modulation. Cluster size = 63 voxels. Coronal, Sagittal and Axial views. 56.8% of cluster in right Area 6 (x = +8, y = –2, z = +55; MNI +8/+2/+50). 41.8% of cluster in left Area 6 (x = +0, y = +4, z = +65; MNI +0/+8/+60).

4.1. Activation of primary auditory cortex

Beginning with the strongest results, the **bilateral modulation of PAC**, it seems more than obvious that listeners' attention increases significantly when FPs/UPs appear in the speech. Top down regulation of primary cortices, e.g. the PAC, has previously been reported (Ghatan et al., 1998) and that heightened attention influences auditory perception has also been shown (Petkov et al. (2004). This 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). However, since UPs *also* modulated PAC in the listeners, conceivably *any* break in the speech signal might serve as a potential attention-heightener. Consequently, the shorter reaction times reported by Fox Tree might also be observed for unfilled pauses or other types of disfluency.

4.2. Activation of motor areas

Perhaps more interesting is the observation that **FPs activate BA6/SMA** in the listening brain. The most obvious explanation for this activation is that when hearing the speaker produce FPs, the listeners prepare to start speaking themselves, i.e. take over the floor. It has been 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). Moreover, an interesting result was reported in Wise et al. (1991) where subjects in a PET study were instructed to silently (i.e. non-vocalizing) generate verbs, which resulted in activation of the SMA, very much in accordance with our own observations. However, it could conceivably be the case that motor cortex activation during speech tasks at least partly occur as a part of motor planning of speech *breathing* (as distinct from baseline breathing), as is pointed out in Murphy et al. (1997).

4.3. Implications for two FP hypotheses

Our observed FP-induced activation of SMA could be seen in the light of the **"floor-holding" hypothesis** of FPs, as first proposed by Maclay and Osgood (1959). This hypothesis suggests that FPs are used by speakers who want to maintain the floor in conversation, as a (semi-deliberate) means to prevent interlocutors from breaking in. Although this might be true, our observations that FPs "kick-start" the speech production system in the listener would indicate that this use of FPs is counter-productive in that the effect on the listener is exactly the opposite of the suggested function to *prevent* interlocutors from speaking, not *preparing* them to do so.

An alternative hypothesis concerning the roles FPs might play in human dialog could be called the **"help-me-out" hypothesis**, as suggested in Clark and Wilkes-Gibbs (1986). This hypothesis suggests that FPs can be used as a (semi-deliberate) signal asking for interlocutor help in that they signal to the listener that the speaker is encountering problems in the production of speech, or simply put: when a speaker is looking for a word or term which is not immediately available to them, uttering "uh" signals to the listener that some help is desired. Our observation that motor areas are activated by FPs should mean that a helpful interlocutor would be faster to come to the rescue.

5. CONCLUSIONS

Three things make this study unique (we believe):

1. We used *fMRI to study disfluency perception*, unlike other perception studies that have relied on EEG and the concomitant 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*.

Our results suggest that FPs – unlike FS and UPs – activate motor areas in the listening brain. However, both FPs and UPs activate PAC, which lends support to the attention-heightening hypothesis that has been forwarded in the literature. 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.

6. ETHICS APPROVAL

The study was approved by the Karolinska Institute ethics committee on April 4, 2007.

7. REFERENCES

Alario, F.-X., Chainay, H., Lehericy, S., Cohen, L. 2006. The role of the supplementary motor area (SMA) in word production. *Brain Research* 1076, 129–143.

Amunts, K., Schleicher, A., Zilles, K. 2007. Cytoarchitecture of the cerebral cortex—More than localization. *NeuroImage* 37, 1061–1063.

Barr, D. , Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes* 25(4), 441–455.

Brickner, R.M. 1940. A Human Cortical Area Producing Repetitive Phenomena When Stimulated. *Journal of Neurophysiology* 3, 128–130.

Clark, H.H., Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22, 1–39.

Corley, M., MacGregor, L.J., Donaldson, D.I. 2007. It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition* 105, 658–668.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25, 1325–1335.

Eickhoff, S.B., Heim, S., Zilles, K., Amunts, K. 2006. Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *NeuroImage* 32, 570–582.

Eickhoff, S.B., Paus, T. Caspers, S., Grosbras, M.-H., Evans, A.C., Zilles, K., Amunts, K. 2007. Assignment of functional activations to probabilistic areas revisited. *NeuroImage* 36, 511–521.

Eklund, R. 2010. The Effect of Directed and Open Disambiguation Prompts in Authentic Call Center Data on the Frequency and Distribution of Filled Pauses and Possible Implications for Filled Pause Hypotheses and Data Collection Methodology. In: *Proceedings of DiSS-LPSS Joint Workshop 2010,* University of Tokyo, Japan, 25–26 September 2010, Tokyo, Japan, 23–26.

Eklund, R. 2004. *Disfluency in Swedish human–human and human–machine travel booking dialogues.* PhD thesis, Linköping University, Sweden.

Ferreira, F., Bailey, K.G.D. 2004. Disfluencies and human language comprehension. *TRENDS in Cognitive Sciences* 8(5), 231–237.

Fox Tree, J.E. 1995. The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language* 34, 709–738.

Fox Tree, J.E. 2001. Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition* 29(2), 320–236.

Fraundorf, S.H., Watson, D.G. 2011. The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language* 65, 161–175.

Friston, K., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D. (eds.). 2007 (second edition). *Statistical Parametric Mapping.* Amsterdam: Elsevier.

Genovese, C.R., Lazar, N.A., Nichols, T. 2002. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate, *NeuroImage* 15, 870–878.

Ghatan, P.H., Hsieh, J.C., Petersson, K.-M., Stone-Elander, S., Ingvar, M. 1998. Coexistence of Attention-Based Facilitation and Inhibition in the Human Cortex. *Neuroimage* 7(1), 23–29.

Goldberg, G. 1985. Supplementary motor area structure and function: Review and hypotheses. The *Behavioral and Brain Sciences* 8, 567–616

Iacoboni, M. 2008. The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology – Paris* 102, 31–34.

Maclay, H., Osgood, C.E. 1959. Hesitation Phenomena in Spontaneous English Speech. *Word* 5, 19–44.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., Zilles, K. 2001. Human Primary Auditory Cortex: Cytoarchitectonic Subdivisions and Mapping into a Spatial Reference System. *NeuroImage* 13, 684–701.

Murphy, K., Corfield, D.R., Guz, A., Fink, G.R., Wise, R.J.S., Harrison, J., Adams, L. 1997. Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology* 83(5), 1438–1447.

Oldfield, R.C. 1971. The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia* 9, 97–113.

Petkov, C.I., Kang, X., Alho, K., Bertrand, O., E. Yund, E.W., Woods, D. L. 2004. Attentional modulation of human auditory cortex. *Nature Neuroscience* 7(6), 658–663.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.-J., Zilles, K. 2001. Probabilistic Mapping and Volume Measurement of Human Primary Auditory Cortex. *NeuroImage* 13, 669–683.

Reich, S.S. 1980. Significance of Pauses for Speech Perception. *Journal of Psycholinguistic Research* 9(4), 379–389.

Wilson, S., Saygın, A.P., Sereno, M.I., Iacoboni, M. 2004. Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* 7(7), 701–702.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., Frackowiak, R. 1991. Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain 114*, 1803–1817.