INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

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, Sweden
[2] Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden
[3] Osher Center for Integrative Medicine, Karolinska Institute, Stockholm, Sweden

robert.eklund@liu.se, martin.ingvar@ki.se

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.
Index Terms: 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 well-known characteristic of spontaneous spoken language is that almost no one is completely fluent but exhibits several types of disfluency, i.e. phenomena like pauses (silent and "filled"), segment prolongations, truncations and so on. The two most common types of disfluency are unfilled pauses (UPs), i.e. silences, and filled pauses (FPs), "eh" and "ehm". The average frequency of disfluency has been reported to be around 6% at word-level [1,2,3,4,5] while the reported average frequency of FPs ranges from 1.9% to 7.6% [6]. (Note that UP frequency is harder to estimate since UPs are more difficult to identify unambiguously.)

Speech disfluency has been studied from at least since the 1930s (for an overview, see [5: pp. 51–172]), and while it is common to regard speech disfluencies simply as "errors" in speech production, there are several studies that indicate that certain kinds of disfluencies can have beneficial effects on listener perception [1,7,8,9,10,11]. While several behavioral studies of speech disfluency have been carried out, few neurocognitive studies have been performed, and most of these studies have used electrophysiological methods and stimuli based on scripted or enacted disfluencies (e.g. [12]). Also, these studies have focused on the effect of speech disfluency on higher-level speech processing – like syntactic parsing – rather than study the effect of disfluencies *proper*. The present study uses functional Magnetic Resonance Imaging (fMRI) to analyze the effect of authentic disfluencies proper to study the effect of unfilled and filled pauses on brain processing.

2. Data collection and method

2.1. Stimulus data

The stimulus data used were excerpts from the spontaneously produced human–human dialog speech data described in Eklund [5: pp. 187–190], and consisted of travel booking dialogs, collected in a Wizard-of-Oz setting.

Subjects were asked to play the role of travel agents listening to customers making travel bookings over the telephone, following a task sheet which provided instructions in mainly iconic ways, so as to not provide the subjects with predefined verbal biases [5:185].

From the original stimulus 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 and distribution of UPs and FPs in spontaneous speech.

2.2. Subjects

The subjects were 16 healthy adults (9F/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 [13]. No specific cut-off value was applied since all subjects were solidly right-handed. 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 fee was also administered.

2.3. 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).

2.4. Experimental design: event-related fMRI

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

http://dx.doi.org/10.21437/Interspeech.2016-973

Stimulus data, with interstimulus times, are shown in Table 1.

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

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 adhesive 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 reported 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 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 × 1 × 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 × 3.75 × 4 mm^3).

In total, T2*-weighted images were collected from 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 [14]. 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 (three translational and three 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 [15,16,17,18].

3. Analysis 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

Given the pioneering character of our study, a whole brain analysis was used.

The results were calculated with a False Discovery Rate (FDR) at $p < 0.05$ [19] with a cluster level threshold of 10 contiguous voxels. Results are shown in Table 2.

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

For **FP > FS**, increased activation was found in Primary Auditory Cortex (PAC) [20,21] 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. Typical modulation is shown in Figure 2.

For **FP > UP**, activation was very similar to that of FPs over FS. Typical modulation is shown in Figure 3.

The results suggest that FPs and UPs equally affect attention (i.e. PAC) in the listener, but while UPs modulate 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, FS 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

4.1. Activation of primary auditory cortex

Beginning with the strongest results, the **bilateral modulation of PAC**, it seems likely 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 [22], and that heightened attention influences auditory perception has also been shown [23]. 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 [1,10]. 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 1944 [24] that SMA is active in the processing of speech, and several later studies have confirmed that both SMA and pre-SMA play a role in both speech *production* [25,26] as well as speech *perception* [27,28]. Furthermore, an interesting result was reported in [29] where subjects in a PET study were instructed to silently (i.e. non-vocalizing) generate verbs, which resulted in activation of the SMA. 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 [30].

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 [14] and the SPM Anatomy Toolbox [16,17,18]. Brodmann Areas were identified using the Talairach Atlas [31] 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

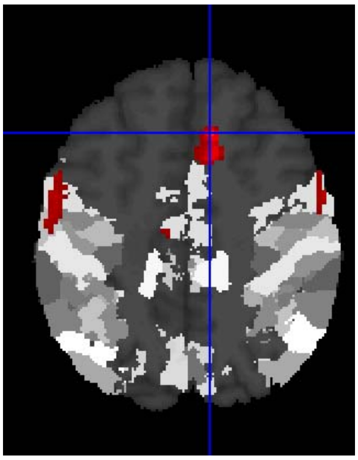

Figure 1. FPs > FS. 101 voxels. 10% of cluster in right area 6. For all three figures: Axial view; Right = right hemisphere. Activation color scheme (increased) red–yellow–white.

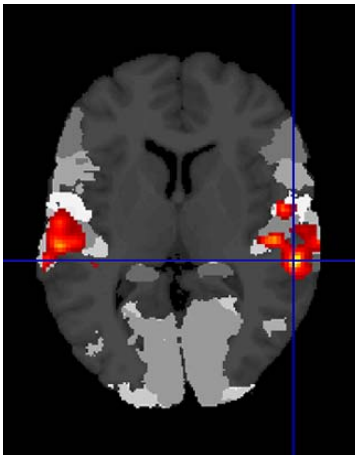

Figure 2. UPs > FS. 1499 voxels. 10.3% of cluster in right TE 3. 9.2% of cluster in right TE 1.0. 5.9% of cluster in right OP 1. 5.6% of cluster in right TE 1.2.

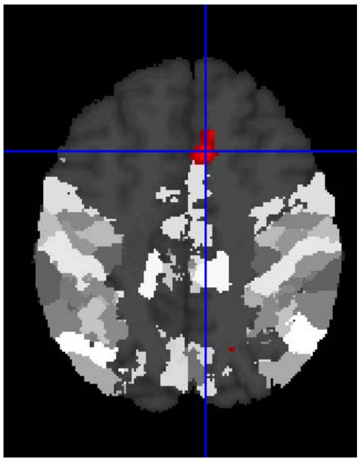

Figure 3. FPs > UPs. 60 voxels. 13.1% of cluster in right area 6.

4.3. Implications for two popular 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 in 1959 by Maclay and Osgood [32]. This hypothesis suggests that FPs are used (semi-deliberately) by speakers in order to hold the floor in conversation, preventing interlocutors from breaking in. While this might well 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 is what could be called the "help-me-out" hypothesis [33] which suggests that FPs can be used (semi-deliberately) as a signal asking interlocutors for help when the speaker is having problems in producing speech and need interlocutor help. If indeed motor areas are activated by FPs, a helpful interlocutor would be faster to come to the rescue.

5. Conclusions

The present study is interesting for three reasons:

1. We used *fMRI to study disfluency perception*, unlike other 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, admittedly 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.

6. Acknowledgements

The study was approved by the Karolinska Institute ethics committee on April 4, 2007. Data were made available by formal agreement between TeliaSonera, Karolinska Institute and the first author. Deep thanks to Peter Fransson for many insightful comments on the design and analysis of this study. Also thanks to Örjan de Manzano (né Blom) and Katarina Gospic for nice scanning assistance and company.

7. References

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

[2] S. Oviatt. Predicting spoken disfluencies during human–computer interaction. *Computer Speech and Language*, vol. 9, pp. 19–35, 1995.

[3] S.E. Brennan, M.F. Schober. How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, vol. 44, pp. 274–296, 2001.

[4] K. Bortfeld, S.D. Leon, J.E. Bloom, M.F. Schober, S.E. Brennan. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, vol. 44, no. 2, pp. 123–147, 2001.

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

[6] R. Eklund. 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. *Proceedings of DiSS-LPSS Joint Workshop 2010,* University of Tokyo, 25–26 September 2010, Tokyo, Japan, pp. 23–26, 2010.

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

[8] D. Barr, M. Seyfeddinipur. The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes,* vol. 25, no. 4, pp. 441–455, 2010.

[9] F. Ferreira, K.G.D. Bailey. Disfluencies and human language comprehension. *TRENDS in Cognitive Sciences*, vol. 8, no. 5), pp. 231–237, 2004.

[10] J.E. Fox Tree. Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition,* vol. 29, no. 2, pp. 320–236, 2001.

[11] S.S. Reich. Significance of Pauses for Speech Perception. *Journal of Psycholinguistic Research*, vol. 9, no. 4, pp. 379–389, 1980.

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

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

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

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

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

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

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

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

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

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

[22] P.H. Ghatan, J.C. Hsieh, K.-M. Petersson, S. Stone-Elander, M. Ingvar. Coexistence of Attention-Based Facilitation and Inhibition in the Human Cortex. *Neuroimage*, vol. 7, no. 1, pp. 23–29, 1998.

[23] C.I. Petkov, X. Kang, K. Alho, O. Bertrand, E.W. Yund, D.L. Woods. Attentional modulation of human auditory cortex. *Nature Neuroscience*, vol. 7, no. 6, pp. 658–663, 2004.

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

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

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

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

[28] S. Wilson, A.P. Saygın, M.I. Sereno, M. Iacoboni. Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, vol. 7, no. 7, pp. 701–702, 2004.

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

[30] K. Murphy, D.R. Corfield, A. Guz, G.R. Fink, R.J.S. Wise, K. Harrison, L. Adams. Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology* vol. 83, no. 5, pp. 1438–1447, 1997.

[31] J. Talairach, P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers, 1988.

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

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