

따라 말하기 과제에서의 음향적 처리와 음운적 처리

유세진¹ (idenia@snu.ac.kr), 이경민^{1,2} (lee.kyoungmin@gmail.com)

¹서울대학교 협동과정 인지과학 전공, ²서울대학교 신경과

Acoustic and phonological processes in the repetition tasks

Sejin Yoo¹ and Kyoung-Min Lee^{1,2}

¹Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Korea

²Department of Neurology, Seoul National University, Seoul, Korea

Abstract

Speech shares acoustic features with other sound-based processing, which makes it difficult to distinguish phonological process from acoustic process in speech processing. In this study, we examined the difference between acoustic process and phonological process during repetition tasks. By contrasting various stimuli in different lengths, we localized neural correlates of acoustic process within bilateral superior temporal gyrus, which was consistent with the previous studies. The activated patterns were widely overlapped between words and pseudowords, i.e., contents-free. In contrast, phonological process showed left-lateralized activation in middle temporal gyrus located at anterior temporal areas. It implies that phonological process is contents-specific as shown in our previous study, and at the same time, more language-specific. Thus, we suggest that phonological process is distinguished from acoustic process in that it is always accompanied with the obligatory access to available phonological codes, which can be an entry of the mental lexicon.

INTRODUCTION

In auditory processing point of view, speech and nonspeech share many underlying processes (Koelsch et al., 2009). According to the dual stream model (Hickok and Poeppel, 2004; Hickok and Poeppel, 2007), phonological process may be distinguished from acoustic process. Furthermore, phonological grammar in a language can shape neural responses to same acoustic signals (Jacquemot et al., 2003). Anatomically, both superior and middle temporal areas are known for spectrotemporal and phonological analysis of auditory signals. More specifically, the transition from perceptual to lexical-semantic analysis of auditory sounds emerges at N400m response, i.e., 200-300ms after speech onset, within superior temporal gyrus (Bonte et al., 2005). That is, there is probably a functional divergence of speech processing in terms of acoustic and phonological view.

In this vein, repetition task, as a fundamental process of speech, is suitable for contrasting acoustic and phonological processes. When we repeat immediately what we hear, a whole language loop, i.e., a speech-motor loop and sound-meaning interface, is formulated around perisylvian region. In repetition task, therefore, we can examine how language *in situ* is processed in different stimuli. Additional advantage when we use repetition tasks is that no assumption is required to study neural activities. Usually, two different tasks are designed and contrasted to show the difference between two cognitive processes. However, the repetition task simply requires same things, i.e., repeating the sounds, for all conditions.

During the tasks, phonological short-term memory or verbal working memory would be recruited (Jacquemot and Scott, 2006; Gruber et al., 2007) and we contrasted stimuli in different lengths. In this way, we aimed to localize neural activities that were linearly correlated to speech analysis in both spectrotemporal and phonological ways. In addition, we presuppose that phonological analysis is more linguistic one and is distinguished from acoustic analysis. Thus, to clarify phonological effects of the stimuli, we adjusted acoustic features of each stimulus by adding irrelevant sounds and contrasted them again.

The result shows acoustic process is localized within specific neural areas, bilateral superior temporal gyrus, irrespective of their specific contents. In contrast, phonological process is more lateralized in left temporal areas and seems to be contents-specific, which inevitably activates subsequent semantic process.

MATERIALS AND METHODS

Subjects

Twelve Koreans (5 males and 7 females; aged 19-34 years, averaged 23.42 years) were recruited as volunteers in experiment 1, after giving an informed consent. Each subject had a normal auditory ability, with no history of neurological disorders. Their handedness was assessed by Edinburgh Test (Oldfield, 1971), which showed all subjects were right-handed. In experiment 2, eleven Koreans (4 males and 7 females; aged 19-34 years,

averaged 24.64 years) in different group were recruited as volunteers. All subjects also had no history of neurological disorders. Their handedness and auditory ability were well assessed and were suitable for the experiment.

Task and Stimuli

With an event-related experimental paradigm, the subjects had two seconds of listening to stimuli. After listening, another two seconds for repeating them was given. That is, the subjects should repeat immediately what they heard (*repetition tasks*). For this 4s of stimuli-responses (S-R) epoch, no EPI images from scanner were generated to reduce the noise level during the tasks.⁵ Thus, the subjects could listen and repeat the stimuli in a comfortable state. Following S-R epoch, an EPI epoch lasted for 12s which generated 6 EPI images. Consequently, 16s were taken for one trial and one test session consisted of 12s of dummy scanning and 21 trials (Fig. 1).

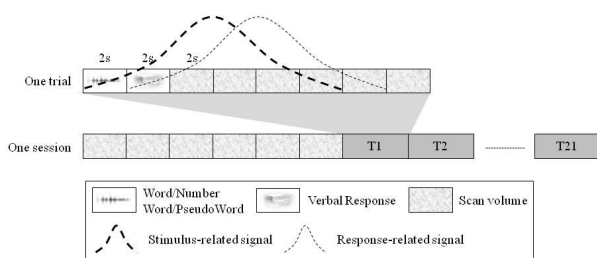


Fig. 1. Experimental paradigm

In the experiment 1, *words* and *numbers* with variable lengths (1, 4, or 7 syllables/digits) were presented to the subjects in random order during one session. For 348s of total session, the subjects were instructed to repeat the stimuli overtly only during response period. Each subject accomplished 6 scans, three for *words* and three for *numbers*. The length of each stimulus and the order of scan conditions were balanced by randomizing them between subjects.

In the experiment 2, *words* and *pseudowords* in fixed lengths (7 syllables) were introduced to make each stimulus have same acoustic energy. By adding acoustic noise into the original stimuli, we generated equally-sized stimuli in length. Except for the contents of stimuli, all tasks were same with the experiment 1. The subjects were instructed to repeat the stimuli only, irrespective of the irrelevant sounds. The irrelevant sounds were made by reversing acoustic signal of syllable [a] temporally. By doing this, all stimuli had same acoustic energy in fixed lengths. Each subject had 6 scans, three for *words* and three for *pseudowords*.

Data Acquisition

⁵ Refer to the following Data Acquisition clause for detail protocols.

To enhance the subject's audibility during the scanning, we introduced sparse sampling method, *interleaved silent steady state* (ISSS) imaging (Schwarzbauer et al., 2006). Compared to *continuous temporal acquisition* or *clustered temporal acquisition*, ISSS has a temporal resolution similar to event-related imaging within a single trial (Gaab et al., 2007; Zaehle et al., 2007; Schmidt et al., 2008). In this way, we could reduce scanner noise during the tasks dramatically. It is also suitable for event-related design in fMRI, which is more sensitive to hemodynamic response for each stimulus.

However, some assumptions about the time-to-peak of the evoked hemodynamic response are required to use ISSS imaging. Furthermore, it reduces the amount of EPI data. Therefore, ISSS imaging avoids the difficulties in choosing, *priori*, the time at which the evoked hemodynamic response will reach its peak, by acquiring multiple volumes. In this study, we assumed hemodynamic response delay of 6s for each trial. Accordingly, we had two different patterns of hemodynamic response, i.e., stimulus-related signal and response-related signal, and it could compensate the lack of task-related signals caused by scan noise (Fig. 1).

In one trial, we had 6 EPI images and totally 132 EPI images were available from one session including initial 6 dummy EPI images. The MR scanner (Siemens Avanto 1.5T) with ISSS sequence was applied in both experiments. T1-weighted anatomical images were obtained first (TR 585ms and TE 7.8ms) and then T2*-weighted EPI images at the same slice locations (132 brain volumes, TR 2000ms, TE 30ms, flip angle 90°, field-of-view 220 x 220 mm², matrix size 64 x 64 pixels, 30 slices with a thickness of 3.4 mm separated by 3.4 mm, parallel to the AC-PC plane) that covered the whole brain except for the cerebellum were obtained.

Data Analysis

SPM5 was used to preprocess and analyze brain image data. All images were realigned, Gaussian-filtered, and spatially normalized to MNI templates. In addition, slice timing of each EPI image was adjusted before statistically analyzed because we introduced event-related design. Based on the general linear model (GLM), each conditional event onset was set at the starting point of S-R epoch. However, it led to unexpected results that hemodynamic response function (HRF) was not matched to the neural response, due to two missing images before task-related signals were evoked. Therefore, we adjusted event onset time by 4s before the original event onset and generated new HRF function.

The individual data were gathered to verify *random effect* in the group data. By statistical inference, the group data were mapped on the template brain in MNI coordinates ($p < 0.001$, unc.). The peaks of activated clusters were summarized in Talairach coordinates (Table 1).

RESULTS

First, we examined the recruited areas during repeating words and numbers in the experiment 1. The results were largely overlapped with the results of our previous study (Yoo et al., 2009), which was around perisylvian region (Fig. 2). The perisylvian region including frontal premotor areas was generally involved in repeating auditory sounds. It included bilateral temporal regions such as superior and middle temporal gyrus with greater activations in posterior parts. More specifically, bilateral superior temporal gyrus (BA41, 42, 22), left transverse temporal gyrus (BA42), bilateral cingulate gyrus (BA23, 24), bilateral precentral gyrus (BA6), right superior frontal gyrus (BA6), bilateral medial frontal gyrus (BA6), and left inferior temporal gyrus had significant activations.

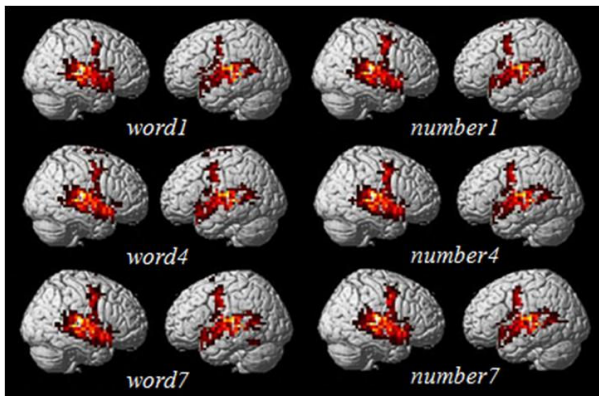


Fig. 2. Activated regions in the experiment 1 ($p < 0.001$, unc.)

The activations shown in the above appear to be common for all conditions, regardless of the length or content of the stimuli. The common activations might indicate 1) *recoding* of auditory stimuli, temporal storage of analyzed sounds, and 3) articulation of auditory stimuli (Henson et al., 2000; Paulesu et al., 1993; Vallar et al., 1997). Unlike block design, however, the event-related design could reveal the subtle difference between each condition by showing event-locked responses (Fig. 3). At activated clusters, we observed different neural responses for each condition as shown in Fig. 3. The response was large for long stimuli, whereas the response was small for short stimuli. Importantly, the responses were locked at the event onset. In other words, neural responses when the subjects repeated what they heard were modulated by the length of stimuli even if all other conditions were the same.

Next, we localized neural activities positively or negatively modulated by the length of stimuli. For words, left superior frontal areas were negatively correlated with the length of stimuli. For numbers, in contrast, bilateral middle frontal areas and right anterior cingulate were negatively correlated with the length of stimuli. Overall,

the frontal areas showed negative correlation with the length of stimuli (Fig. 4a).

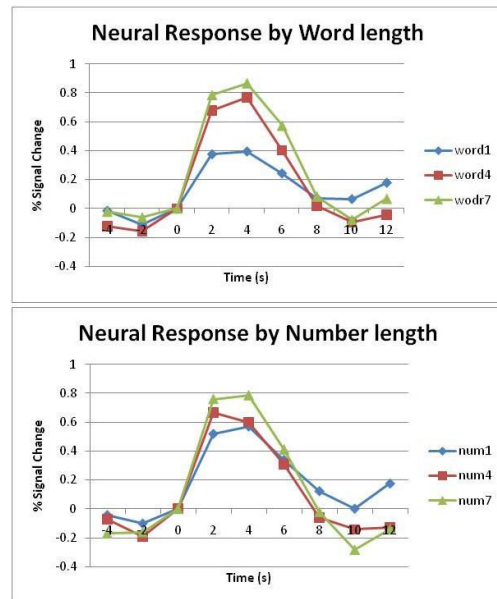


Fig. 3. Neural responses by stimuli length

Conversely, the temporal areas showed positive correlation with the length of stimuli: bilateral superior temporal gyrus and transverse temporal gyrus (Fig. 4b). Interestingly, activations for words and numbers were mostly overlapped within superior temporal gyrus, in contradistinction to negative correlation.

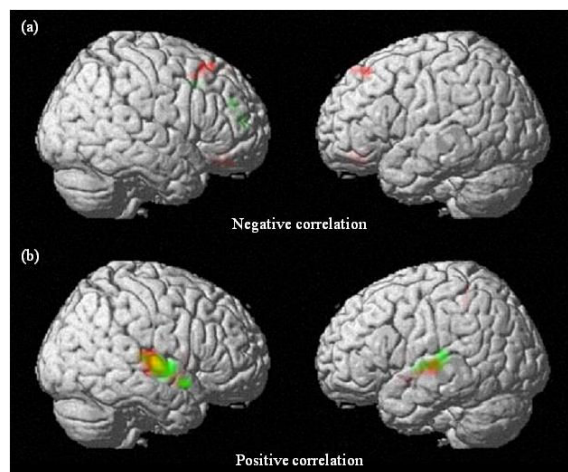


Fig. 4. Neural activities modulated by the length of stimuli ($p < 0.001$, unc.): (a) negative correlation (b) positive correlation, words (green) and numbers (red).

Assuming that other conditions were the same, positive correlation with the length of stimuli shows that superior temporal areas are sensitive to the length of auditory stimuli, regardless of their specific contents. Thus, it might be related to spectrotemporal analysis (Hickok and Poeppel, 2004) and more sensitive to the acoustic energy of auditory stimuli, not to phonological features. In this view, we examined neural activities

evoked by *pure phonological processing* of the stimuli with acoustic energy balanced. In experiment 2, we contrasted words and pseudowords in same lengths to show phonological difference between both stimuli. For words, we found left middle temporal gyrus (BA21) near anterior temporal lobe and parahippocampal gyrus were positively correlated with the length stimuli. For pseudowords, in contrast, there was no positive correlation. We summarized the local maxima of clusters which were positively correlated with the length of stimuli in Table 1.

Table 1. Local maxima of each activated cluster

Brain Regions	Cluster Size	x	y	z	peak t
<i>Word7 > Word4 > Word1 (Experiment 1)</i>					
R. Superior Temporal Gyrus	105	51	-12	0	10.28
L. Transverse Temporal Gyrus	58	-44	-27	12	6.89
<i>Num7 > Num4 > Num1 (Experiment 1)</i>					
R. Superior Temporal Gyrus (BA42)	85	59	-23	12	7.73
L. Parietal Lobe	13	-23	-40	50	7.33
L. Sub-lobar	27	-24	7	-7	6.74
L. Superior Temporal Gyrus (BA22)	36	-59	-16	1	5.57
R. Sub-lobar	28	12	4	3	5.48
<i>Word7 > Word4 > Word1 (Experiment 2)</i>					
L. Middle Temporal Gyrus (BA21)	27	-55	-5	-17	8.55
L. Parahippocampal Gyrus	15	-28	-36	-12	7.69

Only activation with a $p < 0.001$ (unc.) and a volume of at least 640mm^3 (10 measured voxels) were considered. The x , y , z values show the center of gravity of the activated cluster in Talairach coordinates. L., left; R., right; BA, Brodmann area.

Lastly, we mapped neural activities modulated by the stimuli in same length and different length. The results are shown in Fig. 5. As described, two neural activities had different spatial signatures in the temporal lobe.

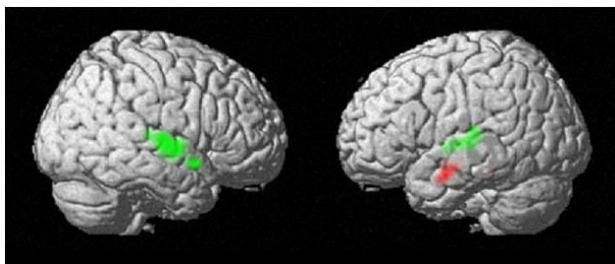


Fig. 5. Neural activities positively correlated by the length of words ($p < 0.001$, unc.): results from experiment 1 with different length of stimuli (green) and experiment 2 with same length of stimuli (red).

When the length of stimuli was linearly varied by the length of words to repeat, the positive correlation was observed in the superior temporal gyrus. In contrast, when the length of stimuli was fixed irrespective of the length of words to repeat, the positive correlation was observed in anterior temporal area, i.e., middle temporal gyrus. More notably, while the former was observed bilaterally, the latter was observed only in left temporal

area. It probably indicates that the latter might activate language-specific areas in contrast to the former.

DISCUSSION

Verbal language uses acoustic sounds and thus phonological processing in speech is intrinsically intermingled with acoustic processing. This in part results from that human brain utilizes similar neural substrates for dealing with various sounds in different categories (Koelsch et al., 2009). Exactly the same thing happens during repetition tasks. The repetition tasks that are based on speech-motor loop automatically access to acoustic and phonological processes in the same phonological loop. According to previous studies, i.e., *recall* tasks with lexical and phonological distracters, the phonological store seems to operate at the level of individual phonemes or syllables rather than at lexical levels. In other words, sounds appear to be represented at phonological level before manipulated.

Once sounds are represented phonologically in repetition tasks, the incoming sounds have obligatory access to the entry of corresponding mental lexicon. However, if it fails to find phonological entries available, phonological output codes are made and articulated without additional semantic process (Howard and Nickels, 2005). The semantic process is mainly determined by specific contents of the sounds. There is also a contents-specificity, which leads to a divergence of phonological process (Yoo et al., 2009). In contrast, acoustic process only concerns the physical characteristics of auditory sounds itself and thus is likely to be restricted within specific processing area. Specifically, acoustic analysis of words is carried out in the superior temporal cortex (Price, 2000; Hickok and Poeppel, 2007).

By contrasting auditory stimuli in different lengths, in the first experiment, we localized acoustic process within bilateral superior temporal gyrus as parts of phonological short-term memory systems (Fig. 4b). The fact that other regions showed no correlation with the length of stimuli indicates that acoustic process is positioned in the early steps of speech processing. At this step, auditory sounds are contents-free, that is, words and numbers have same activation patterns. That is, the auditory sounds are analyzed spectrotemporally before phonological input codes are accessed. Note that it showed no laterality. It implies that the observed activation was acoustic, not phonological. Interestingly, we found some frontal areas were negatively correlated with the length of stimuli (Fig. 4a). These were distinct for each stimulus in contradistinction to positive correlation and might be correlated to verbal working memory operations.

Notably, however, the linearly modulated activation was not observed in case the subjects repeat acoustically balanced stimuli. Instead, left middle temporal gyrus at anterior temporal lobe was significantly activated (Fig.

5). In addition, the activation was not found in case the subjects repeat pseudowords, i.e., it would be language-specific process. Therefore, given that the stimuli have same acoustic energy, the activation observed here is correlated with a phonological process, not an acoustic process. In this case, the absence of activation in superior temporal gyrus is due to that there is no difference between each stimulus in terms of acoustic energy.

The activation in anterior temporal areas was known for *combinatorial process* of complex words at lexical or higher levels of language processing (Humphries et al., 2001). As most Korean words in 7 syllables are complex words that require combinatorial process to build a specific context, it might be automatically recruited during repetition tasks. In other words, phonological processing of speech probably involves spontaneous neural activities at lexical level. With respect to this, there is evidence that language related areas such as Heschl's gyrus, planum temporale, and inferior frontal gyrus can be modulated by high-level linguistic functions, e.g., semantic operations, as well as low-level stimulus queue (Noesselt et al., 2003).

However, top-down modulation of high-level linguistic modules seems to have little influences on this interaction because the subjects have no *a priori* knowledge or instructions about the stimuli before they repeat them. But rather, it would be related to the interaction between multiple processes, operating in parallel during the tasks. Each neural area in the dual-stream model is also interactive, not sequential, in this sense. Thus, we suggest that phonological process is always accompanied with additional semantic process to access to the entry of mental lexicon. In contrast, acoustic process has no such subsequent process because phonological codes are not made yet, in this case.

CONCLUSIONS

In conclusion, we localized acoustic process in bilateral superior temporal gyrus. Also, phonological process is more lateralized in left temporal areas and at the same time, it inevitably activates subsequent semantic process. Both processes are integrated to process speech sounds in cooperative way and the obligatory access to available phonological codes can be an entry of the mental lexicon.

REFERENCES

1. Baddeley, A.D., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158-173.
2. Bonte, M., Parviainen, T., Hytönen, K., Salmelin, R. (2005). Time Course of Top-down and Bottom-up Influences on Syllable Processing in the Auditory Cortex. *Cerebral Cortex*, 16, 115-123.
3. Gaab, N., Gabrieli, J.D.E., and Glover, G.H. (2007). Assessing the Influence of Scanner Background Noise on Auditory Processing. I. An fMRI Study Comparing Three Experimental Designs with Varying Degrees of Scanner Noise. *Human Brain Mapping*, 28, 703-720.
4. Gruber, O., Müller, T., and Falkai, P. (2007). Dynamic interactions between neural systems underlying different components of verbal working memory. *Journal of Neural Transmission*, 114, 1047-1050.
5. Henson, R.N.A., Burgess, N., and Frith, C.D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, 38, 426-440.
6. Hickok, G. and Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92, 67-99.
7. Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature*, 8, 393-402.
8. Howard, D. and Nickels, L. (2005). Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology*, 22:1, 42-77.
9. Humphries, C., Willard, K., Buchsbaum, B., and Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *NeuroReport*, 12, 1749-1752.
10. Jacquemot, C. and Scott, S.K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Science*, 10:11, 480-486.
11. Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological Grammar Shapes the Auditory Cortex: A Functional Magnetic Resonance Imaging Study. *J. of Neuroscience*, 23(29), 9541-9546.
12. Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., and Gruber, O. (2009). Functional Architecture of Verbal and Tonal Working Memory: An fMRI Study. *Human Brain Mapping*, 30, 859-873.
13. Noesselt, T., Jon Shar, N., and Jäncke, L. (2003). Top-down and bottom-up modulation of language related areas – An fMRI Study. *BMC Neuroscience*, 1471-2202/4, 1-12.
14. Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9:1, 97-113.
15. Paulesu, E., Frith, C.D., and Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-345.
16. Price, C.J. (2000). Functional imaging studies of aphasia. In Mazziotta J.C., Toga, A.W., Frackowiak, R.S.J. (Eds), *Brain mapping: the disorders*. San Diego: Academic Press. 181-200.
17. Schmidt, C.F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., and Jancke, L. (2008). Silent and Continuous fMRI Scanning Differentially Modulate Activation in an Auditory Language Comprehension Task. *Human Brain Mapping*, 29, 46-56.
18. Schwarzbauer, C., Davis, M.H., Rodd, J.M., and Johnsrude, I. (2006). Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*, 29, 774-782.
19. Vallar, G., Betta, A.M.D, and Silveri, M.C. (1997). The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. *Neuropsychologia*, 35:6, 795-812.

20. Yoo, S., Jeon, H., and Lee, K.M. (2009). Auditory stimuli-specific divergence of phonological map in human cortical areas. *Proceedings of the Korean Society for Cognitive Science*, 88-94, May 15-16, Seoul, Korea.
21. Zaehle, T., Schmidt, C.F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P., and Jancke, L. (2007). Comparison of “silent” clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *NeuroImage*, 37, 1195-1204.