



HAL
open science

Somatosensory influence on auditory cortical response of self-generated sound

Nozomi Endo, Coriandre Emmanuel Vilain, Kimitaka Nakazawa, Takayuki Ito

► **To cite this version:**

Nozomi Endo, Coriandre Emmanuel Vilain, Kimitaka Nakazawa, Takayuki Ito. Somatosensory influence on auditory cortical response of self-generated sound. *Neuropsychologia*, 2025, 211 (May), pp.109103. <10.1016/j.neuropsychologia.2025.109103>. <hal-04980015>

HAL Id: hal-04980015

<https://hal.science/hal-04980015v1>

Submitted on 6 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Title: Somatosensory influence on auditory cortical response of self-generated sound

Nozomi Endo¹, Coriandre Vilain², Kimitaka Nakazawa¹, Takayuki Ito²

¹ Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1, Komaba, Meguro-ku, Tokyo 153-8902, Japan

² Univ. Grenoble Alpes, Grenoble-INP, CNRS, GIPSA-lab, 11 rue des Mathématiques, Grenoble Campus BP46, F-38402 SAINT MARTIN D'HERES CEDEX, France

Corresponding author:

Takayuki Ito, Ph. D.

Univ. Grenoble Alpes, Grenoble-INP, CNRS, GIPSA-lab

Email: takayuki.ito@gipsa-lab.grenoble-inp.fr

1 **Acknowledgements**

2 This study was supported by research grants from Japan Society for the Promotion of
3 Science (JSPS) KAKENHI (20J14743) and the National Institute on Deafness and
4 Other Communication Disorders (R01- DC017439 and R01-DC022097). We would like
5 to thank Dr. David J Ostry and Dr. Jean-Luc Schwartz for helpful comments, and Ms.
6 Elisabeth Anggianto for her support of data recording and preprocessing.

7

8 **Keywords**

9 Electroencephalography, Event-related potentials, Auditory-motor interaction, Auditory
10 attenuation, Multisensory integration

11

12 **Abstract**

13 Motor execution which results in the generation of sounds attenuates the cortical
14 response to these self-generated sounds. This attenuation has been explained as a result
15 of motor relevant processing. The current study shows that corresponding
16 somatosensory inputs can also change the auditory processing of a self-generated sound.
17 We recorded auditory event-related potentials (ERP) in response to self-generated
18 sounds and assessed how the amount of auditory attenuation changed according to the
19 somatosensory inputs. The sound stimuli were generated by a finger movement that
20 pressed on a virtual object, which was produced by a haptic robotic device.
21 Somatosensory inputs were modulated by changing the stiffness of this virtual object
22 (low and high) in an unpredictable manner. For comparison purposes, we carried out the
23 same test with a computer keyboard, which is conventionally used to induce the
24 auditory attenuation of self-generated sound. While N1 and P2 attenuations were clearly
25 induced in the control condition with the keyboard as has been observed in previous
26 studies, when using the robotic device the amplitude of N1 was found to vary according
27 to the stiffness of the virtual object. The amplitude of N1 in the low stiffness condition
28 was similar to that found using the keyboard for the same condition but not in the high
29 stiffness condition. In addition, P2 attenuation did not differ between stiffness

30 conditions. The waveforms of auditory ERP after 200 ms also differed according to the
31 stiffness conditions. The estimated source of N1 attenuation was located in the right
32 parietal area. These results suggest that somatosensory inputs during movement can
33 modify the auditory processing of self-generated sound. The auditory processing of self-
34 generated sound may represent self-referenced processing like an embodied process or
35 an action-perception mechanism.

36

37 **Introduction**

38 In our daily lives, we generate various sounds with our own movements, such as
39 speaking, knocking, or playing a musical instrument. These self-generated sounds are
40 generated by precisely controlling movement in interaction with the external
41 environment. Since the sound generated is the goal of this motor execution, it is
42 important to understand how the self-generated sound is processed during movement.

43 Our own movements are known to affect the auditory processing of the sounds
44 they induce and this is known as auditory attenuation. In studies involving
45 magnetoencephalography (MEG) and electroencephalography (EEG), it is found that
46 the first negative peak (N1) and the second positive peak (P2) of the recorded potentials
47 or fields is smaller than that of externally generated sounds (Aliu et al., 2009; Baess et
48 al., 2009, 2011; Martikainen et al., 2005; Schafer & Marcus, 1973). This auditory
49 attenuation may be due to sensory attenuation during movement in general or, in other
50 terms, a suppression of sensory processing or perception (Chapman et al., 1987; Cohen
51 & Starr, 1987). The possible functional role may also be explained by the idea of
52 “corollary discharge” in biology (Sperry, 1950) or “efference copy” in motor control
53 (Von Holst, 1954). The presence of motor execution which is obtained from efference

54 copy or corollary discharge can be used to distinguish whether a sound was generated
55 by a person's own actions or by someone else, through the gating of sensory information
56 (Blakemore et al., 1998).

57 The attenuation of cortical potentials or fields could also be the result of
58 auditory-motor interaction or integration due to variability of movement. Neszmeélyi &
59 Horváth (2017) showed that the amplitude of auditory ERPs of self-generated sound
60 varied according to the amplitude of the force produced by pinching that is used to
61 produce the sound. A similar variation has also been found in speaking-induced auditory
62 suppression (Niziolek et al., 2013). The cortical responses recorded using
63 magnetoencephalography showed that auditory attenuation induced by speaking was
64 reduced when the produced sound was less prototypical. Moreover, changes in auditory
65 attenuation may not always correspond to changes in sensory suppression. Eliades &
66 Wang (2003) suggested that auditory inhibitory neurons related to auditory attenuation
67 contribute by increasing the dynamic range of the auditory response to the subject's own
68 vocalizations. Indeed, auditory-motor integration can also improve the loudness
69 discrimination of self-generated sound (Endo et al., 2021; Myers et al., 2020). All these
70 authors suggest the involvement of auditory-motor integration in the auditory
71 processing of self-generated sound. However, motor execution involves both motor

72 commands and somatosensory feedback. While motor outputs in the form of efference
73 copy or corollary discharge may explain a part of auditory attenuation function, it is still
74 unclear how somatosensory inputs during movement affect or interact with the auditory
75 processing of self-generated sound.

76 Somatosensory inputs can interact with and be integrated into auditory
77 processing in general (Stein & Meredith, 1993). An electrophysiological study
78 involving macaques showed that somatosensory input is reflected in the responses of
79 auditory cortex neurons (Schroeder et al., 2001). Vibrotactile sensation on the fingertip
80 has also been found to induce activation of the auditory cortex in humans (Caetano &
81 Jousmäki, 2006). At a behavioral level, vibrotactile stimulation systematically biases
82 auditory frequency and intensity perception (Yau et al., 2010). In addition to tactile
83 sensation, somatosensory feedback related to movement also changes auditory
84 processing in case of speech perception. When somatosensory stimuli related to speech
85 movements were externally applied during a speech perception task in the absence of
86 actual movement, the perception of speech sounds was biased towards the sound
87 corresponding to speech movement associated with the somatosensory stimulation (Ito
88 et al., 2009; Trudeau-Fisette et al., 2019). Since this modulation of speech processing is
89 related to body movement, this auditory-somatosensory integration may be part of an

90 embodied process or the use of an action-perception linkage (Ito et al., 2021). Although
91 this interaction has been demonstrated in orofacial system, it still remains unclear how
92 somatosensory inputs during movement more generally affect or interact with auditory
93 processing of self-generated sounds.

94 This study examines whether somatosensory inputs change the auditory
95 cortical response to self-generated sound. To achieve this aim, we set up an original
96 experimental framework enabling the expected somatosensory interference to be
97 modulated while keeping the motor interference fixed. This was based on the reasoning
98 that if the modulation in somatosensory interference modified the auditory response,
99 this would be due to an auditory-somatosensory rather than an auditory-motor effect.
100 We recorded auditory ERPs in response to self-generated sound, in which a pure tone
101 was produced in conjunction with finger tapping movement. In order to modulate
102 somatosensory inputs during the movement, we used a haptic robotic device, in which
103 the stiffness of a virtual object (VO) was varied unpredictably between two different
104 stiffness levels. The important detail of our set-up is that, while the motor task was fixed
105 (tapping on the robotic device), it was associated to varied somatosensory inputs, but
106 the unpredictable presentation made it impossible for the participant to adapt the motor
107 command accordingly, hence ensuring the stability of the motor input. The

108 quantification of the motor/somatosensory input on the auditory response was based on
109 the subtractive method that is classically used in the study of multisensory integration
110 (Calvert, 2001). The principle is to compare the cortical response to the two information
111 sources (here, auditory and motor/somatosensory) simultaneously to the summation of
112 the responses to each individual source. Therefore, we systematically compared
113 auditory-motor ERPs (responses to self-generated sound) with uni-sensory responses in
114 each modality alone (auditory-alone response: response to externally generated sound
115 and motor-alone response: response to movement without sound presentation). We
116 defined the specific auditory response in an auditory-motor-somatosensory task by
117 subtracting the motor-alone ERP from the auditory-motor ERP, as is classically the case
118 in experiments in this field (see e.g. Arnal et al., 2009; Baess et al., 2011; Ito et al.,
119 2014; Neszmeilyi & Horváth, 2017). Then, we compared these specific auditory
120 responses obtained in the two somatosensory conditions of motor interference, both
121 together and with the recorded auditory-alone response. For purposes of comparison, we
122 also recorded auditory ERPs to self-generated sounds produced with a computer
123 keyboard, to assess how motor effects associated to the two somatosensory conditions
124 (two stiffness levels) would compare with possible motor effects associated to a current
125 experimental environment. All comparisons were focused, as is classically the case in

126 experiments in this field (Calvert, 2001), on the two early cortical ERP components that
127 are the first negative (N1) and the first positive peak (P2). Our expectation in the current
128 study was that different somatosensory inputs can change the auditory cortical response
129 to self-generated sound in spite of a fixed motor task. The modulation of auditory-motor
130 interaction would be the result of an interaction of auditory and somatosensory inputs,
131 thus providing a clear answer to the question of the potential role of the somatosensory
132 modality in audio-motor interactions in the motor-induced auditory attenuation effect.

133

134 **Methods**

135 **Participants**

136 Sixteen healthy participants (18–50 y.o. 11 women) with normal hearing and
137 who are right-handed (both factors were self-reported), were tested in the experiment.
138 Sample size was determined based on previous studies of self-generated sounds (Baess
139 et al., 2011, Neszmeélyi & Horváth, 2017, Reznik et al 2015). The participants were
140 naïve to the purposes of the experiment. The protocol of this experiment was approved
141 by the Comité d’Ethique pour la Recherche, Grenoble Alpes (CERGA: Avis-2020-04-
142 28-06). All participants signed the corresponding consent form.

143 **Somatosensory stimulation**

144 A haptic robotic device (Phantom 1.0, SensAble Technologies) was used to
145 produce mechanical perturbations during participants' motor tasks (See Fig. 1A). The
146 device had a hook at the end of the robotic arm. Displacement, velocity, and force at the
147 endpoint of the robotic arm are available in three dimensions. The control and recording
148 were carried out with a sampling frequency of 1000 Hz.

149 The participants placed their right index finger on the hook and were asked to
150 make a finger tapping movement as if they were pressing a computer keyboard. The
151 task involved pressing downward on the robotic arm until they made contact with a soft
152 foam placed 15 mm below the start position, and then returning to their initial posture
153 by releasing the pressing force immediately after the contact.

154 As shown in Fig. 1A, the configuration of the robotic arm was maintained in
155 the servo-control position with a specific stiffness. When participants pressed the
156 robotic arm in the finger tapping task, they received a different reaction force according
157 to the stiffness level of the position controller (inside small panel in Fig. 1A). We
158 applied two different stiffness levels (Low: 0.15 N/mm and High: 0.35 N/mm). In the
159 low stiffness condition, participant received smaller reaction forces than in the high

160 stiffness condition during movement. These two stiffness conditions were tested in one
161 recording session and changed unpredictably so participants could not adapt their motor
162 commands and were likely to keep these commands fixed all through the session,
163 whatever the applied stiffness level. Since the task with haptic robot is like pressing
164 against a virtual object that differs in stiffness (Low versus High), we called the session
165 with the haptic robot a Virtual Object (VO) session.

166 As a control condition to verify the amount of auditory attenuation in the
167 situation comparable with the previous study and to compare with the results of VO
168 session, we also imposed the motor task in pressing a key on the computer keyboard, as
169 a replacement of the robotic device. We used a computer keyboard to maintain
170 compatibility with the previous studies, because it was familiar to the participant, and
171 because it is an easily available device that can be used to obtain the time of pressing for
172 auditory stimulus presentation. Since it is not possible to switch quickly between these
173 two devices, we carried out a separate session for the right index finger tapping (Key
174 session). As mentioned below, we also included this task with the left index finger
175 tapping in both the VO and Key sessions to examine possible session differences in
176 EEG signals, in particular as regards the amount of auditory attenuation.

177 **Auditory stimulation**

178 The auditory stimulus was a 1000 Hz pure tone, 50 ms in duration, which was
179 presented binaurally through EEG-compatible earphones (Model TIP-300, Natus). The
180 stimulus was either presented alone with no finger movement (externally-generated
181 sound condition) or in a self-generated process, where the stimulus was presented in
182 conjunction with the finger tapping movement when the vertical displacement of the
183 robotic arm reached 7 mm below the initial position (VO session), or when a key press
184 was detected in the task with a keyboard (Key session).

185

186

187

188

189

190

191

192 **Procedure**

193 The main goal of this experiment was to examine whether the auditory ERP
194 associated with self-generated sound changed because of the somatosensory input
195 resulting from the movement involved in the production of these sounds. In the main
196 recording session (VO session), we recorded auditory ERPs of self-generated sound
197 (auditory-motor response: AM ERP) under the two somatosensory conditions (two
198 stiffness levels: High and Low). In the same session, we also recorded auditory ERPs of
199 externally generated sound (auditory-alone response: A ERP) and ERPs in response to
200 the movement task alone (motor-alone response: M ERP). In the same way, in the other
201 session with the computer keyboard (Key session), we recorded auditory ERPs of self-
202 generated sound (AM ERP) together with auditory-alone and motor-alone ERPs (A and
203 M ERPs).

204 To smoothly record AM, M and A ERPs, we designed the experimental
205 procedure as follows (Fig. 1B). The participants' task involved alternate finger tapping
206 between the left and right index fingers. As a nominal task, we asked the participants to
207 carry out this alternate finger tapping with a consistent tempo (100 bpm). At the
208 beginning of each trial, the auditory sound stimuli were presented four times at this

209 tempo (100 bpm) in order to provide an example of the tempo of the tapping movement.
210 They also enabled A ERPs to be recorded. Then, participants repeated the tapping
211 sequence four times with each finger, alternating between the left and right fingers at the
212 100 bpm tempo. The finger order (starting with left or right) was counterbalanced using
213 a visual cue every four trials. Still, it should be noted that the motor/somatosensory
214 situation for left-finger tapping was fixed (always keyboard action) while the conditions
215 of interest only concerned the right finger (High vs. Low stiffness in the VO session,
216 keyboard in the Key session).

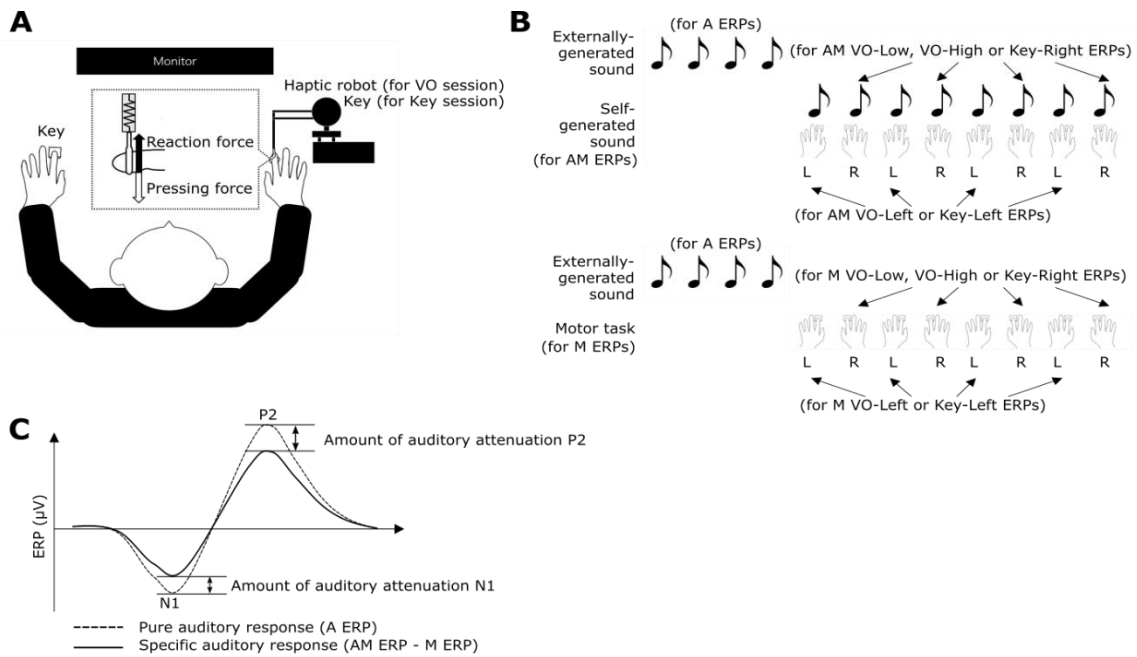
217 In order to make the task movement similar in auditory-motor and motor-alone
218 recordings, the same trial sequence using alternate finger tapping was used for the
219 recording of both AM and M ERPs. When the auditory stimulus was presented in
220 conjunction with the finger tapping (see auditory stimulation section), AM ERPs were
221 recorded, and when the finger tapping was carried out without any sound production, M
222 ERPs were recorded. These two recordings (AM and M ERPs) were carried out in one
223 recording session. Hence, each trial consisted in a sequence of 4 A ERP recordings
224 followed by 8 AM or M ERP recordings with alternate left and right fingers. Importantly,
225 the Low and High stiffness conditions alternated randomly within the 4 right-finger
226 tapping in the VO session, with 2 occurrences of each stiffness condition in each trial.

227 Auditory-motor and motor-alone recordings randomly alternated with 2 of each within
228 sequences of 4 consecutive trials.

229 As two different devices were used, the VO and Key tests were carried out in
230 separate sessions. After one session was completed, the other was followed after a short
231 break. The order of VO and Key sessions was counter-balanced across the participants.
232 Still, the recording of AM and M ERPs during the left index finger tapping (VO-Left
233 and Key-Left) aimed at checking that the two sessions were matched except for the key
234 experimental manipulations, since the task for the left index finger was the same in both
235 VO and Key sessions.

236 In total, we carried out 128 trials (64 each for AM and M ERPs) in the VO
237 session and 64 trials (32 each for AM and M ERPs) for the Key session. In the VO
238 session, each trial consisted of 4 auditory stimuli and 8 motor tasks (2 for VO-Low, 2
239 for VO-High, and 4 for VO-Left). Altogether this provided 512 A ERPs (VO-Audio),
240 256 AM ERPs with the right finger tapping (128 each in the two stiffness conditions:
241 VO-Low and VO-High), 256 AM ERPs with the left finger tapping (VO-Left), 256 M
242 ERPs with the right index finger tapping (128 each in the two stiffness conditions), and
243 256 M ERPs with the left index finger tapping. In the Key session, we recorded 256 A

244 ERPs (Key-Audio), 128 AM ERPs with the right index finger tapping (Key-Right), 128
245 AM ERPs with the left index finger tapping (Key-Left), 128 M ERPs with the right
246 index finger and 128 M ERPs with the left index finger.



247

248 **Fig. 1 A)** The experimental setup in the VO session. The small panel between the hands of the
 249 participant represents the reaction to the pressing forces according to the stiffness level of the
 250 position controller of the haptic robot. The left finger always presses a key (VO-Left or Key-
 251 Left), the right finger presses either a virtual object with varied stiffness (VO-Low or VO-
 252 High) or a key (Key-Right). **B)** The task sequence of one trial with alternate finger tapping and
 253 self-generated sound. For AM ERPs, finger tapping is accompanied by the sound, while it is
 254 not for M ERPs. The trial begins by a sequence of 4 sounds providing A ERPs. **C)** Principle of
 255 the ERP analysis. Peak detection provides N1 and P2 estimation. Then, A ERPs provide pure
 256 auditory response, while combined AM and M ERPs provide specific auditory responses for
 257 each motor situation. In addition, the difference of peaks between pure and specific auditory
 258 responses provides the amount of auditory attenuation.

259 **Table. 1 Condition of interest in Within-Session comparisons.**

		Peak Amplitude (Pure and Specific)						Results		
		VO session			Key session					
		Audio	Low	High	Left	Audio	Right	Left	N1	P2
Within- Session	VO	x	x	x	x				Fig. 3A (Left panel)	Fig. 5A (Left panel)
	Key					x	x	x	Fig. 3A (Right panel)	Fig. 5A (Right panel)

260

261

262 **Table. 2 Condition of interest in Between-Session comparisons.**

		Amount of Attenuation (Specific - Pure)					Results	
		VO session			Key session			
		Low	High	Left	Right	Left	N1	P2
Between- Session	Right	x	x		x		Fig. 3B (Right panel)	Fig. 5B (Right panel)
	Left			x		x	Fig. 3B (Left panel)	Fig. 5B (Left panel)

263

264

265 **EEG acquisition and preprocessing**

266 EEG was recorded using a 64-channel EEG system (Biosemi ActiveTwo
267 System) in 512 Hz sampling rate. The Biosemi system uses CMS/DRL electrodes as
268 ground electrodes for recording. To exclude components of eye blink and movement,
269 electro-oculography was recorded, in which electrodes were placed at the outer corner
270 of both eyes and at the top and bottom of the left eye. We also used a fixation cross-
271 mark to minimize eye blink and movement.

272 EEGLAB (Delorme & Makeig, 2004) was used for preprocessing. EEG signals
273 were filtered offline using a 1–30 Hz bandpass filter and re-referenced to the average
274 across all electrodes (Yao et al., 2019). Independent component analysis (Onton &
275 Makeig, 2006) was applied to exclude components corresponding to noise and artifact,
276 including eye-blink and motion, which were identified by manual inspection.
277 Continuous EEG signals were segmented into epochs between –200 and 400 ms relative
278 to the onset of auditory stimulation. An initial offset was removed from each epoch
279 using the average amplitude in the pre-stimulus interval (–200 to –100 ms). Finally, the
280 processed ERPs were averaged across trials in each type of ERP (A ERP in each

281 condition: VO-Audio and Key-Audio, and AM and M ERPs in each condition: VO-Low,
282 VO-High, VO-Left, Key-Right, and Key-Left) on a per-participant basis.

283 In the data analysis, we first carried out single-electrode analysis to
284 characterize the changes in the ERPs of self-generated sound more effectively,
285 following the traditional approach to auditory ERP analysis and making sure to avoid
286 the issue of multiple comparisons across electrodes. We focused on Fz in the 10-20
287 system based on our previous observation of auditory-somatosensory interaction (Ito et
288 al., 2014) and general knowledge of auditory ERPs that the largest amplitude of
289 auditory ERP is usually concentrated in the frontal region of the midsagittal plane. Peak
290 amplitudes of N1, P2 and ERPs after 200 ms were extracted using a 30 ms time window
291 centered on the local maximum peak in the grand average signal, or in other terms, the
292 average of all conditions in one session (see grey vertical bars in Fig. 2 and Fig. 6). We
293 then applied source localization using all 64 electrodes at the point where a difference
294 was found in the single-electrode analysis.

295

296

297

298 **Statistical analyses of ERP components**

299 With reference to the wealth of literature on multisensory interaction (Calvert
300 2001), we evaluated the auditory-motor interaction based on the assumption that the
301 recorded AM ERPs would equal the linear summation of the A ERPs and M ERPs if no
302 auditory-motor interaction were involved. From this basis, we applied the framework of
303 auditory attenuation as in previous studies (Baess et al., 2011; Ito et al., 2014;
304 Neszemélyi & Horváth, 2017), by subtracting the M ERP from the AM ERP in a given
305 condition (specific auditory response, see Fig. 1C). The subtraction of the M ERP also
306 eliminates the possibility that the observed ERP change was due to differences in the
307 produced movement. We then examined if this specific auditory response is different
308 from the pure auditory responses obtained by A ERP (Audio in the following), and
309 compared specific auditory responses among given conditions. These were, for the VO
310 session, low and high stiffness for the right finger (Low and High in the following) and
311 keyboard tapping for the left finger (Left in the following), and for the Key session, left
312 and right finger tapping (Left and Right in the following) (see Fig. 1C).

313 The first and crucial questions for this study were whether the ERPs changed
314 between the two stiffness conditions and how ERPs obtained with the stiffness

315 manipulation differed from those with conventional key-pressing task. In the analysis of
316 N1 and P2 peaks, we first carried out Within-Session comparisons to verify whether and
317 how pure auditory response was different from the specific auditory responses in each
318 session (see Table. 1). We compared the peak amplitudes directly in each session (Audio,
319 Low, High and Left in the VO session and Audio, Right and Left in the Key session),
320 and applied one-way repeated-measures ANOVA with the post-hoc test using the
321 Shaffer procedure (Shaffer, 1986).

322 We then carried out Between-Session comparisons using the amount of
323 auditory attenuation, in which pure auditory response (A ERP) were subtracted from
324 specific auditory responses (AM ERP - M ERP) (see Table. 2). This comparison was
325 carried out for each finger, i.e. VO-Low, VO-High, and Key-Right for the right finger
326 and VO-Left and Key-Left for the left finger, as different cortical areas are involved in
327 the processing of each finger movement. The analysis of the right finger is the main
328 analysis to examine whether the amount of auditory attenuation differed according to
329 the stiffness conditions, and whether and how those are different from results obtained
330 for the conventional key-pressing task. Analysis of the left finger may provide possible
331 session differences between VO and Key in terms of auditory attenuation, as the left
332 finger tapping task was carried out under the same conditions as the VO and Key

333 sessions described above. We applied one-way repeated-measures ANOVA with the
334 post-hoc test using the Shaffer procedure in each finger analysis. In the analysis of the
335 left finger, we also applied a two-way ANOVA with the factors of sessions (VO and
336 Key) and auditory responses (pure and specific) for a more precise comparison to find a
337 baseline difference between sessions. It should be noted that we did not apply two-way
338 ANOVA in the analysis of the right finger, as different levels are involved in the
339 auditory response factor during the VO session (one pure auditory response and two
340 specific auditory responses: High and Low).

341 In addition to the analysis of N1 and P2, we also compared ERPs after 200 ms.
342 It is possible that different experimental devices induce different temporal patterns of
343 ERPs in this period, so we focused only on the difference between the two stiffness
344 conditions (VO-High and VO-Low) and did not consider the contrasts with Left and
345 Audio conditions. As shown in the Results (Fig. 2), we observed that most of the
346 difference between the two stiffness conditions occurred around the period between 200
347 and 400 ms. We focused on two local peaks within this period, determined based on the
348 grand-average response (222 and 336 ms latency in Fig. 6A). The peak amplitude was
349 calculated using a 30 ms time window about the grand average and compared between

350 the two stiffness conditions using a one-way repeated-measures ANOVA with the post-
351 hoc test using the Shaffer procedure.

352 We also calculated the Bayes factor (BF_{10}) for all comparison of ANOVA.

353

354

355 **Source localization**

356 We applied source localization using the sLORETA/eLORETA software
357 package (Pascual-Marqui, 2002), as in our previous study (Ito et al., 2021). This
358 estimation used a realistic head model (Fuchs et al., 2002) from the Montreal
359 Neurological Institute's MNI152 template (Mazziotta et al., 2001) with the standard
360 electrode positions on the MNI152 scalp corresponding to the Biosemi EEG system.
361 The sLoreta images were calculated for each participant and each auditory ERP
362 comparison. Using these images, we carried out statistical parametrical mapping by
363 applying non-parametrical tests across participants (Nichols & Holmes, 2002), based on
364 estimating the empirical probability distribution for the max-statistic (e.g., the
365 maximum of a t-stat) under the null hypothesis via randomization. A log F-value was

366 applied to compare ERPs between pure auditory response and specific auditory
367 response in each condition. This methodology corrects for multiple electrodes and
368 voxels. We focused on sources related to N1 peaks since possible sources related to the
369 P2 and ERPs following P2 were not observed. In this case, negative values indicate
370 suppression relative to baseline.

371

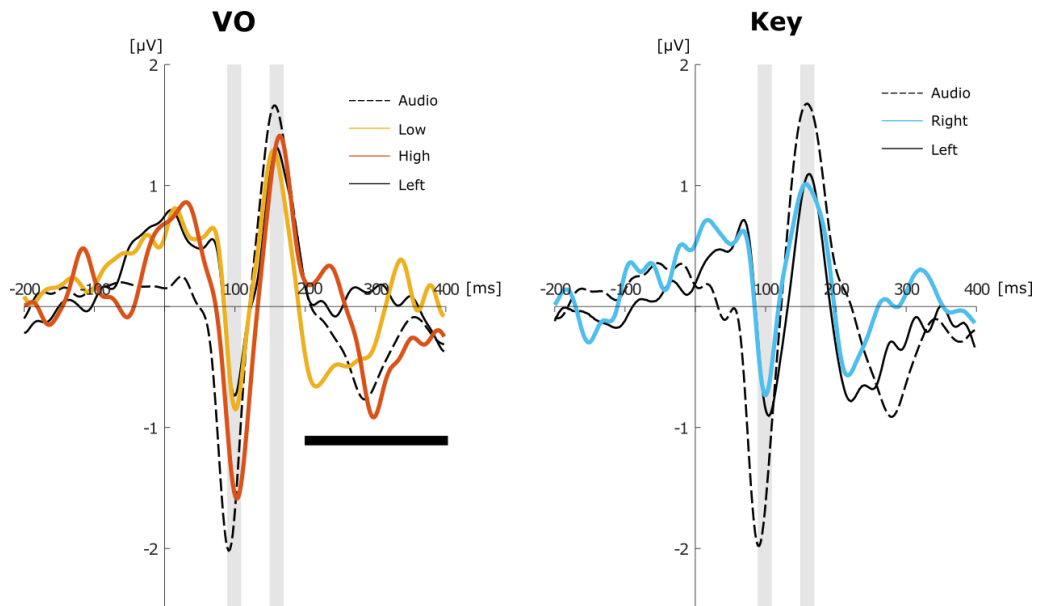
372

373 **Results**

374 Fig. 2 shows the grand average of ERPs at Fz for all conditions. Pure auditory
375 response (A ERPs) showed a conventional pattern with an N1 peak at around 99 ms and
376 a P2 peak at around 159 ms. The temporal pattern of specific auditory responses, in
377 which the M ERP was subtracted from the recorded AM ERPs, also showed N1 and P2
378 peaks similar to those for pure auditory responses. As expected, the peak amplitudes of
379 N1 in the specific auditory responses were significantly less than those in the pure
380 auditory response. This attenuation was also seen at P2 peaks. By comparing two
381 stiffness conditions (VO-High and VO-Low), there was a difference at the N1 peak but
382 not at the P2 peak. The temporal patterns were also different, mostly around the period

383 between 200 and 400 ms represented by the horizontal black bar in the left panel of

384 Fig.2.



385

386 **Fig. 2 Grand averaged ERPs at Fz in the VO (left panel) and Key sessions (right panel). Audio**

387 **corresponds to A ERP. Low, High and Left in VO and Right and Left in Key represent specific**

388 **auditory responses (where M ERPs were subtracted from AM ERPs). Gray vertical bars**

389 **represent a 30 ms time window about the peak amplitudes at N1 and P2. The black horizontal**

390 **bar represents the time period focusing on the difference between Low and High condition**

391 **(VO-Low and VO-High) after 200 ms based on our visual inspection.**

392

393

394 *N1 peak*: We first carried out Within-Session comparisons by evaluating the peak
395 amplitudes in each recording session (Fig. 3A). In the VO session, the one-way ANOVA
396 found significant differences across conditions ($F(3,45) = 16.22, p < 0.001, \eta p^2 = 0.52,$
397 $BF_{10} = 41688.73$). Post-hoc comparisons showed that the amplitude of the N1 peak in
398 the Low and Left conditions was significantly less than in Audio condition ($t(15) = 5.31,$
399 $p < 0.001, d = 0.78, BF_{10} = 323.37$ in Low, $t(15) = 6.79, p < 0.001, d = 0.87, BF_{10} =$
400 3440.73 in Left) and the High condition ($t(15) = 3.09, p = 0.022, d = 0.57, BF_{10} = 6.98$
401 in Low, $t(15) = 4.05, p = 0.003, d = 0.66, BF_{10} = 36.59$ in Left). There was no
402 significant difference between the High and Audio conditions ($t(15) = 1.44, p = 0.339, d$
403 $= 0.21, BF_{10} = 0.61$) or between the Low and Left conditions ($t(15) = 0.75, p = 0.467, d$
404 $= 0.09, BF_{10} = 0.32$). That indicates that typical N1 attenuation was induced by the key-
405 pressing task (Left) and that this N1 attenuation varied according to the magnitude of
406 somatosensory input (Low and High) such that it was reduced during interaction with a
407 relatively hard object (High).

408 In the Key session in which both index fingers tapped a keyboard we found a
409 significant difference across the conditions ($F(2,30) = 8.21, p = 0.001, \eta p^2 = 0.35, BF_{10}$
410 $= 23.03$). Multiple comparisons showed that N1 amplitudes in the Left and Right

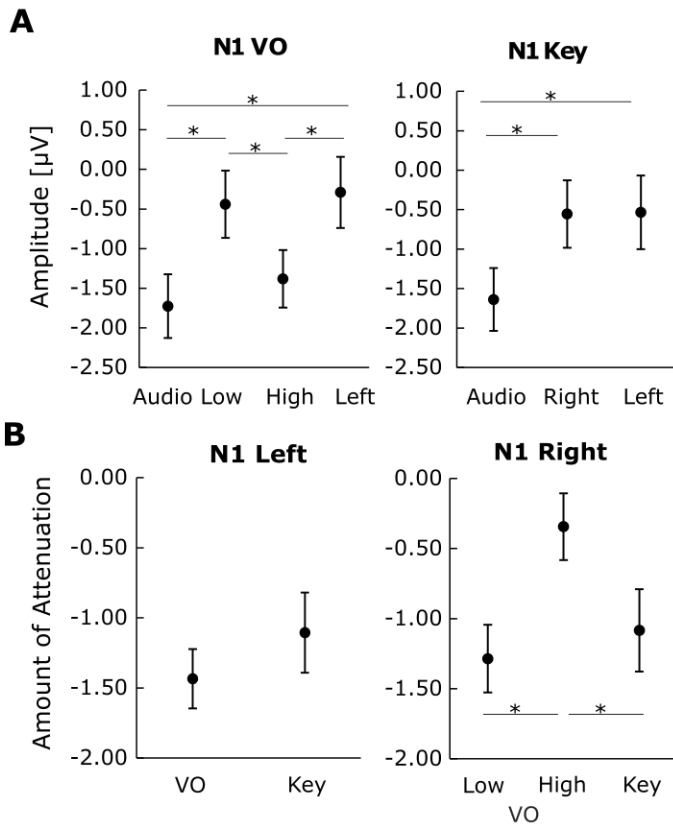
411 conditions were significantly less than in Audio condition ($t(15) = 3.87, p = 0.002, d =$
412 $0.64, BF_{10} = 26.86$ in Left, $t(15) = 3.69, p = 0.002, d = 0.63, BF_{10} = 19.46$ in Right) and
413 that there was no significant difference between the Left and Right conditions ($t(15) =$
414 $0.06, p = 0.952, d = 0.01, BF_{10} = 0.26$). This indicates that the typical N1 attenuation
415 was induced similarly with both hands, which suggests that ERPs associated with key
416 tapping with the left finger can be used for direct comparison in the VO session as if the
417 ERP response had actually been collected using the right finger.

418 We also compared ERPs in the ipsilateral side (Between-Session comparison).
419 For this comparison, we focused on the amount of auditory attenuation, which is the
420 difference found from results for the Audio condition. The right panel in Fig. 3B shows
421 a N1 difference between the right index finger conditions (VO-Low, VO-High and Key-
422 Right). The difference was confirmed by a one-way repeated measures ANOVA
423 ($F(2,30) = 5.64, p = 0.008, \eta p^2 = 0.27, BF_{10} = 7.35$). Multiple comparisons showed that
424 N1 differences in the Key-Right condition were similar to those in the VO-Low ($t(15) =$
425 $0.65, p = 0.523, d = 0.19, BF_{10} = 0.31$), but greater than in the VO-High condition ($t(15)$
426 $= 2.72, p = 0.016, d = 0.71, BF_{10} = 3.77$). As shown in the Within-Session comparison,
427 the N1 differences in the VO-Low were significantly greater than in the VO-High
428 condition ($t(15) = 3.09, p = 0.007, d = 0.91, BF_{10} = 6.98$). The left panel of Fig. 3B

429 shows the response associated with the left index finger. There was no significant
430 difference ($F(1,15) = 1.89$, $p = 0.190$, $\eta p^2 = 0.11$, $BF_{10} = 0.68$). In addition, we also
431 applied a two-way ANOVA with the factor of session (VO and Key) and auditory
432 response (pure and specific). We did not find any significant interaction between session
433 and auditory response ($F(1,15) = 1.90$, $p = 0.190$, $\eta p^2 = 0.11$, $BF_{10} = 0.93$) and main
434 effect of session ($F(1,15) = 0.50$, $p = 0.489$, $\eta p^2 = 0.03$, $BF_{10} = 0.26$), while there was
435 significant main effect of auditory response ($F(1,15) = 33.10$, $p < 0.001$, $\eta p^2 = 0.69$,
436 $BF_{10} = 10.00$). These results indicated that the session did not affect the recorded
437 responses and, more specifically, the amount of auditory attenuation and pure auditory
438 responses.

439 Possible cortical sources of this N1 modulation were estimated by sLoreta. This
440 estimation was carried out in each session separately and involved a comparison with
441 the Audio condition. Significant differences were found mainly in the right parietal and
442 frontal areas at the 5 % significance level. In the VO-High, VO-Key, and Key-Left
443 conditions, statistically significant current sources were found in the parietal area (VO-
444 High [45, -35, 60], VO-Left [60, -50, 40], Key-Left [10, -45, 70] in MNI, see Fig. 4A).
445 Significant sources were also found in the right frontal area for these two conditions
446 (VO-Low [50, 45, -10], Key-Right [5, 40, 55] in MNI). VO-Low and Key-Right

447 conditions also showed significance in the right parietal region (VO-Low [50, -30, 55],
448 Key-Right [10, -50, 70] in MNI, see Fig. 4B). These results suggest that the right
449 parietal region plays a role in the processing of self-generated sound.



450

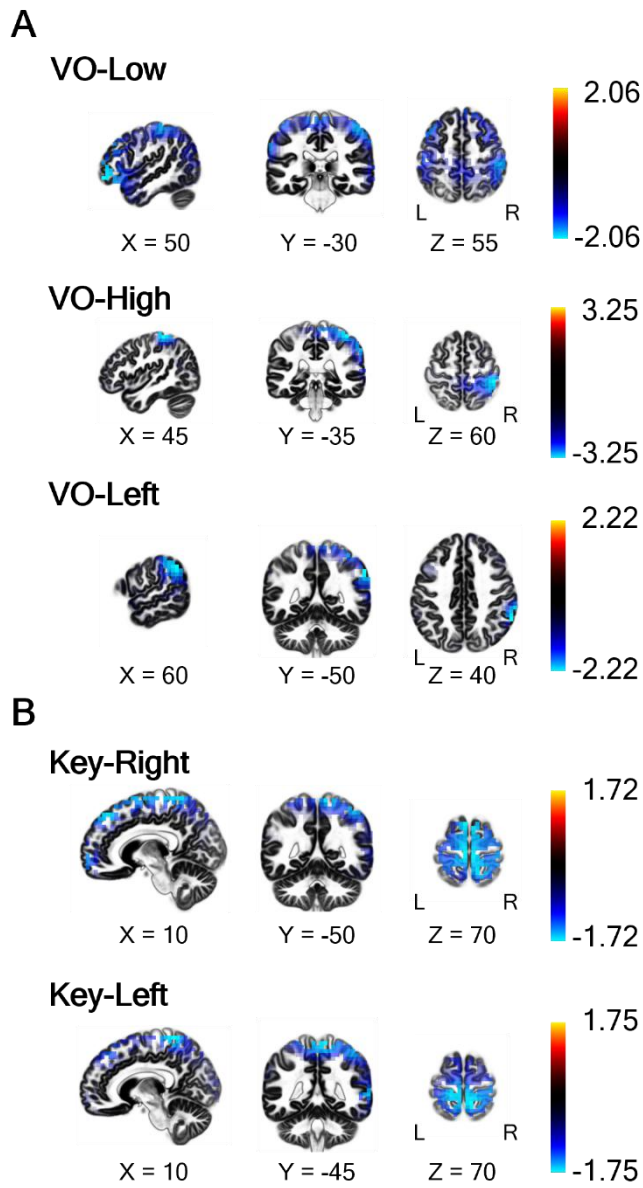
451 **Fig. 3 A) Averaged amplitude of the N1 peak across conditions in each session. B) Averaged N1**

452 **attenuation in each finger. Error bars represent the standard error across participants in both**

453 **panels. The values in the figure represent mean values and standard errors. The asterisk (*)**

454 **represents $p < 0.05$.**

455



456

457 **Fig. 4** The estimated cortical source corresponding to N1 attenuation in the VO (A) and Key

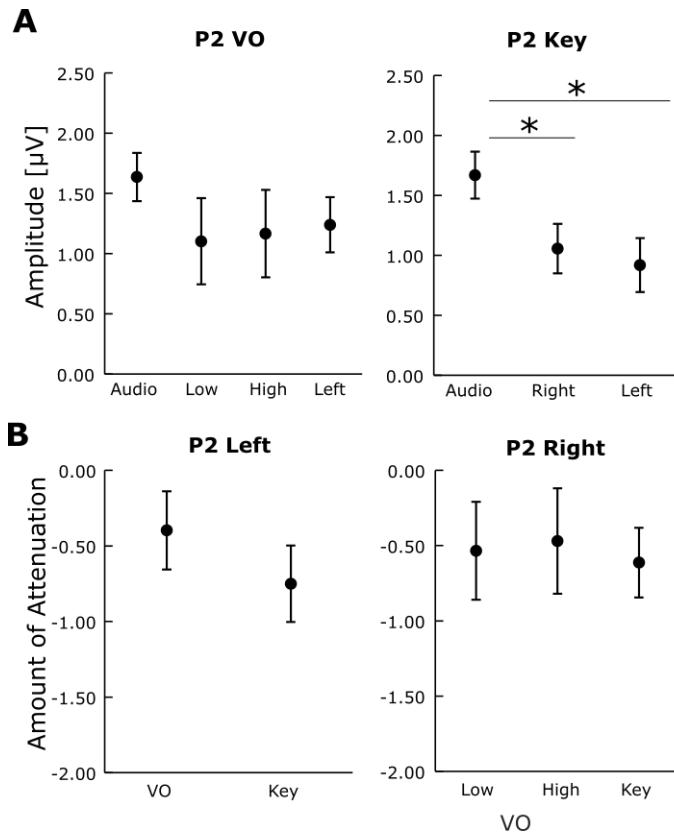
458 sessions (B). Color bars represent the range of sLORETA voxel values.

459

460 *P2 peak*: In Within-Session comparisons, there were no significant differences in the
461 VO session ($F(3,45) = 1.22, p = 0.315, \eta p^2 = 0.08, BF_{10} = 0.29$), but a significant
462 difference was observed in the Key session ($F(2,30) = 6.76, p = 0.004, \eta p^2 = 0.31, BF_{10}$
463 $= 12.60$, see Fig. 5A). Multiple comparisons in the Key session showed that the
464 amplitudes in both Left and Right conditions were significantly less than in the Audio
465 condition ($t(15) = 2.96, p = 0.010, d = 0.90, BF_{10} = 5.63$ in Left, $t(15) = 2.65, p = 0.018,$
466 $d = 0.73, BF_{10} = 3.36$ in Right) and that there was no significant difference between the
467 Left and Right conditions ($t(15) = 0.89, p = 0.390, d = 0.16, BF_{10} = 0.36$). However,
468 when this test was repeated in the context of Between-Session comparisons, no
469 significant difference in between fingers was observed (Right: $F(2,30) = 0.16, p = 0.85,$
470 $\eta p^2 = 0.01, BF_{10} = 0.18$ and Left: $F(1,15) = 1.88, p = 0.191, \eta p^2 = 0.11, BF_{10} = 0.69$ see
471 Fig. 5B). We also applied a two-way ANOVA for the left finger tapping task with the
472 factors of session (VO and Key) and auditory response (pure and specific). We did not
473 find any significant interaction between session and auditory response ($F(1,15) = 1.88,$
474 $p = 0.191, \eta p^2 = 0.11, BF_{10} = 0.69$) and main effect of session ($F(1,15) = 1.03, p = 0.326,$
475 $\eta p^2 = 0.06, BF_{10} = 0.31$), while we found significant main effect of auditory response
476 ($F(1,15) = 6.73, p = 0.020, \eta p^2 = 0.31, BF_{10} = 5.23$). Additional Bayes factor analysis
477 also showed BF_{10} was less than 1 in the case of no significant difference. Accordingly,

478 the results of Between-session comparisons found the amount of P2 attenuation to be
479 similar in the VO and Key sessions and between two stiffness conditions, while the
480 Within-Session comparisons did not show statistical difference from the pure auditory
481 response in the VO session. The latter may be due to the fact that usage of an unfamiliar
482 robotic device causes relatively large variability in VO-High and Vo-Low conditions.
483 The results may indicate that the P2 attenuation is induced consistently across all
484 conditions of self-generated sound although the amount can be varied according to the
485 device for finger tapping task. Like with the result of N1 peak, the session effect
486 between VO and Key sessions may be negligible since we did not find any difference in
487 N1 and P2 peaks between them about the left finger (VO-Left and Key-Left).

488



489

490 **Fig. 5 A) Averaged amplitude of P2 peak across conditions in each session. B) Averaged P2**

491 **attenuation in each finger. Error bars represent the standard error across participants in both**

492 **panels. The values in the figure represent mean values and standard errors. The asterisk (*)**

493 **represents $p < 0.05$.**

494

495

496 *ERPs in later periods (after the P2 peak)*: The focus here was on differences between
497 stiffness conditions (VO-Low and VO-High). Fig. 6A shows a difference between the
498 VO-Low and VO-High conditions. The peaks of this difference were at 226 ms and 333
499 ms (Fig. 6A). At each peak, a one-way repeated measures ANOVA showed a significant
500 difference ($F(1,15) = 8.37, p = 0.011, \eta p^2 = 0.36, BF_{10} = 5.83$ at 226 ms and $F(1,15) =$
501 $4.64, p = 0.048, \eta p^2 = 0.24, BF_{10} = 1.79$ at 333 ms), indicating differences in the cortical
502 response in the VO-Low and VO-High conditions. These results suggest that different
503 somatosensory inputs also change the cortical processing of self-generated sound in the
504 period later than P2.

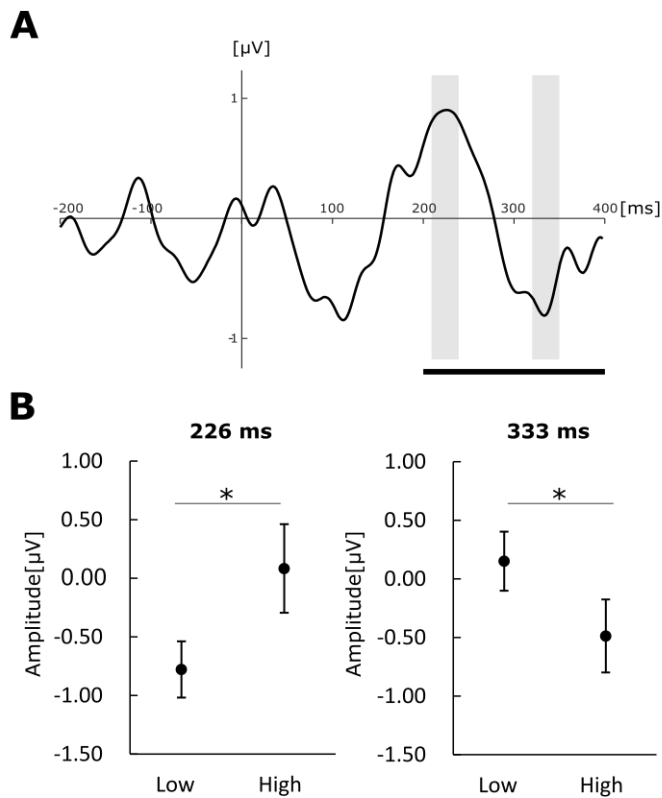
505

506

507

508

509



510

511 **Fig. 6 A) Difference between VO-Low and VO-High. The black horizontal line represents the**

512 **period during which a difference is observed in the amplitude of the VO-Low and VO-High**

513 **responses. The vertical grey bars show 30 ms windows centered at 226 ms and 333 ms for peak**

514 **amplitude calculation. B) Peak amplitude at 226 ms (left panel) and at 333 ms (right panel).**

515 **Error bars represent the standard error across participants. The values in the figure represent**

516 **mean values and standard errors. The asterisk (*) represents $p < 0.05$.**

517

518

519 **Discussion**

520 The purpose of this study was to investigate whether the auditory attenuation of
521 self-generated sound varied according to somatosensory input during the movement that
522 produced the sound generation. We modified somatosensory inputs using a haptic
523 robotic device by unpredictably altering the stiffness of a virtual object. We found that
524 different stiffnesses of virtual objects changed the amount of auditory attenuation in the
525 N1, but not in the P2 peak of the auditory event related potentials. The peak amplitude
526 of N1 in the low stiffness condition was reduced in response to externally generated
527 sound, while the N1 peak in high stiffness condition was almost similar to the one in
528 response to externally generated sound. A difference between two stiffness conditions
529 was also found in the period after the P2 response (200–400 ms). These results suggest
530 that auditory responses to self-generated sound vary according to somatosensory inputs
531 associated with the movements involved in sound generation. We also examined
532 whether the observed auditory responses during interactions with a virtual object (VO)
533 were similar to those observed in conventional computer keyboard tasks (Key). In the
534 Key session, we successfully reproduced the conventional pattern of auditory
535 attenuation in both the N1 and P2 responses. In addition, N1 attenuation in the Key

536 session was similar to that in low stiffness condition of VO session, but different from
537 the one in high stiffness condition. P2 attenuation was similar in Key and VO. This
538 suggests that auditory-somatosensory interactions may work differently in these
539 conditions involving different devices. Finally, the source localization analyses found
540 that the right parietal region may be a possible source of the auditory-somatosensory
541 interaction associated with N1 attenuation, suggesting that the current change may not
542 be due to a gating of auditory processing but instead to additional activity by auditory-
543 somatosensory interaction. Overall, the results demonstrate an auditory-somatosensory
544 interaction in the auditory processing of self-generated sound and likewise suggest that
545 somatosensory inputs on their own may play an important role in the processing of self-
546 generated sound.

547

548 **Somatosensory effect on the auditory attenuation of self-generated sound**

549 As shown in previous studies (Baess et al., 2011; Neszmeélyi & Horváth, 2017),
550 the auditory responses were attenuated when sound was generated by a subject's own
551 movements (self-generated sound). A possible mechanism could be sensory suppression
552 based on inputs associated with motor commands, such as efference copies (Blakemore

553 et al., 1998). Based on this idea, sensory suppression can be induced similarly when the
554 executing motor commands are similar. Contrary to this expectation, we found that the
555 ERP amplitude corresponding to the N1 peak and in the period between 200 and 400 ms
556 were both changed according to the stiffness conditions for the similar finger tapping
557 task. Since the current manipulation involved unpredictable changes in the stiffness of a
558 virtual object, the participants were not able to change or adjust their strategy of motor
559 execution in advance for each stiffness condition. This means that the motor command
560 was similar for these stiffness conditions. Receiving somatosensory inputs were
561 different according to the stiffness conditions, and somatosensory inputs may thus be a
562 main source of the current modulation of ERPs in response to self-generated sounds.
563 The results suggest that somatosensory inputs during movement are also involved in the
564 auditory processing of self-generated sounds. Moreover, the same effects are even
565 observed at short latencies (N1). This auditory-somatosensory link can work in parallel
566 with the previously proposed mechanism based on inputs arising from motor commands.
567 It also raises the possibility that sensory suppression associated with movement is based
568 on sensory feedback rather than efference copy.

569 The differences between stiffness conditions were also seen in two different
570 periods (the N1 peak and the period between 200 ms and 400 ms). N1 is known to be

571 involved in low-level processing such as feature detection related to frequency and
572 intensity of sound (Näätänen & Picton, 1987; Näätänen & Winkler, 1999). On the other
573 hand, the response after 200 ms is likely to be involved in the higher level processing,
574 such as decision making (P300), object-spatial recognition (Demiral et al., 2012), and
575 stimulus discrimination (N200, Patel & Azzam, 2005). The somatosensory inputs may
576 thus affect at least two different phases of auditory processing.

577 In addition, when we compared the amount of auditory attenuation on N1
578 between VO and Key conditions, the auditory attenuation in the low stiffness condition
579 (VO-Low) was found to be similar to that found for the Key-Right condition but not for
580 the high stiffness condition (VO-High). This may be related to the similarity of the
581 somatosensory inputs. The somatosensory inputs in the VO-Low condition can be more
582 similar to the one in the Key-Right condition than the VO-High condition because of
583 smaller reaction force in VO-Low condition. A change in somatosensory inputs due to
584 the greater reaction force in the High Stiffness condition may trigger an additional or
585 distinct auditory-somatosensory interaction process. This process could also play a role
586 in other contexts of auditory-somatosensory interaction (Stein & Meredith, 1993; Yau et
587 al., 2009; Ito et al., 2009). The involvement of general auditory-somatosensory
588 interaction process may be supported by our result of source localization in N1 that the

589 right parietal area may be involved consistently regardless of the left and right finger
590 tapping. This additional process can be modified according to somatosensory inputs,
591 and hence can be seen as the reduction of the amount of auditory attenuation. Further
592 investigation is required to study the possible mechanism involved.

593

594 **Consideration from previously observed auditory attenuation of self-generated**
595 **sound**

596 Auditory attenuation of self-generated sound has been previously observed in N1 and
597 P2 peaks using a key pressing with a computer keyboard (Han et al., 2021; Klaffehn et
598 al., 2019, Knolle et al., 2012, 2013, Tremblay et al., 2014). The current study
599 successfully replicated this attenuation in the Key session. The previous studies showed
600 that different processes may be involved in the attenuation of N1 and P2. N1 attenuation
601 may reflect the temporal coincidence of auditory input and motor output (Han et al.,
602 2021; Klaffehn et al., 2019) and predictions based on motor execution in the cerebellar
603 process (Knolle et al., 2012, 2013). P2 may reflect learning of the auditory-motor
604 interaction in speech (Tremblay et al., 2014) and predict the timing of sound
605 presentation (Klaffehn et al., 2019), and higher-order cognitive functions in motor

606 execution (Knolle et al., 2013). Our results also showed that N1 and P2 were separately
607 modulated between two manipulated conditions (Key and VO). The N1 amplitude was
608 modulated according to the stiffness condition. While the attenuation in the low stiffness
609 condition was that found with the computer keyboard, it was almost removed in the
610 high stiffness condition. On the other hand, the P2 was not different between two
611 stiffness conditions and the amplitude reduction was similar in the VO and Key sessions.
612 These results suggest the involvement of auditory-somatosensory interaction in the
613 auditory processing of self-generated sounds may be different in N1 and P2.

614

615 **Possible cortical origin of auditory-somatosensory processing of self-generated** 616 **sounds**

617 In the current auditory-somatosensory interaction for the self-generated sound
618 processing, a possible cortical source was estimated in the right hemisphere parietal
619 region. This was consistently found for both left and right finger tapping. Previous
620 studies have showed that the inferior parietal area is a center of multisensory integration
621 for the processing of both speech sounds (Hickok & Poeppel, 2007) and non-speech
622 sounds (Tachibana et al., 2010). The parietal cortex is also involved in visuo-motor

623 coordination (Rizzolatti et al., 1997), and may likewise participate in the current
624 auditory-motor interaction for the perception of self-generated sounds. Although our
625 findings suggest a source in the superior parietal lobule, this may be due to difference of
626 somatotopy, in which the sensorimotor cortical area for the limb is located medial to the
627 orofacial area. On the other hand, the consistent observation of the right parietal area's
628 involvement regardless of whether the task involved left and right finger movement
629 may be difficult to reconcile with the contralateral innervation of both the motor and
630 primary somatosensory cortex. Nevertheless, the secondary somatosensory cortex has
631 bilateral innervation (Disbrow et al., 2003), and thus somatosensory inputs from the
632 hand may affect the parietal region of both hemispheres. Conventionally, the left
633 hemisphere is involved in speech auditory processing (Hickok & Poeppel, 2007) and the
634 right hemisphere is in non-speech auditory processing (Brancucci et al., 2005). As the
635 current auditory stimuli involved non-speech sounds, the right parietal area may be
636 dominant in the current auditory and somatosensory integration.

637

638

639

640 **Limitation**

641 In the current data analysis, we mainly compared ERPs on a specific electrode
642 (Fz) and specific time point (N1 and P2) based on prior knowledge concerning the
643 auditory attenuation of self-generated sound, while we used data from all electrodes for
644 source localization analysis. This limitation of electrode and time can be applied
645 typically in order to avoid potential problem of EEG signals (low Signal-to-Noise ratio,
646 relatively small effect size and multiple comparison problem). For future study, an
647 exhaustive search using all electrodes of whole brain and all time points can be applied
648 for better characterization of the current somatosensory effect in ERP.

649

650 **Conclusion**

651 The current study showed the auditory attenuation of self-generated sounds
652 using a finger tapping task with a robotic device and also with a keyboard. The N1 peak
653 and the response in the period between 200 and 400 ms were both modulated according
654 to somatosensory inputs. The attenuation patterns were different for the N1 and P2
655 peaks according to whether the robotic device or a keyboard was used. A possible
656 cortical source of the N1 attenuation was located in the right parietal cortex which is

657 known for its role in multisensory integration or transformation of sensory coordination.

658 The results support the idea that an auditory-somatosensory interaction can be involved

659 in the auditory processing of self-generated sound.

660

661 **References**

- 662 Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced Suppression of the
663 Auditory Cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802.
664 <https://doi.org/10.1162/jocn.2009.21055>
- 665 Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of
666 visual facilitation in speech processing. *Journal of Neuroscience*, *29*(43), 13445-
667 13453. <https://doi.org/10.1523/JNEUROSCI.3194-09.2009>
- 668 Baess, P., Horváth, J., Jacobsen, T., & Schröger, E. (2011). Selective suppression of self-
669 initiated sounds in an auditory stream: An ERP study. *Psychophysiology*, *48*(9),
670 1276–1283. <https://doi.org/10.1111/j.1469-8986.2011.01196.x>
- 671 Baess, P., Widmann, A., Roye, A., Schröger, E., & Jacobsen, T. (2009). Attenuated
672 human auditory middle latency response and evoked 40-Hz response to self-
673 initiated sounds. *European Journal of Neuroscience*, *29*(7), 1514–1521.
674 <https://doi.org/10.1111/j.1460-9568.2009.06683.x>
- 675 Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-
676 produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640.
677 <https://doi.org/10.1038/2870>

678 Brancucci, A., Babiloni, C., Rossini, P. M., & Romani, G. L. (2005). Right hemisphere
679 specialization for intensity discrimination of musical and speech sounds.
680 *Neuropsychologia*, 43(13), 1916–1923.
681 <https://doi.org/10.1016/j.neuropsychologia.2005.03.005>

682 Calvert G. A. (2001). Crossmodal processing in the human brain: insights from
683 functional neuroimaging studies. *Cerebral cortex (New York, N.Y. : 1991)*,
684 11(12), 1110–1123. <https://doi.org/10.1093/cercor/11.12.1110>

685 Caetano, G., & Jousmäki, V. (2006). Evidence of vibrotactile input to human auditory
686 cortex. *NeuroImage*, 29(1), 15–28.
687 <https://doi.org/10.1016/j.neuroimage.2005.07.023>

688 Chapman, C. E., Bushnell, M. C., Miron, D., Duncan, G. H., & Lund, J. P. (1987).
689 Sensory perception during movement in man. *Experimental Brain Research*,
690 68(3), 516–524. <https://doi.org/10.1007/BF00249795>

691 Cohen, L. G., & Starr, A. (1987). Localization, timing and specificity of gating of
692 somatosensory evoked potentials during active movement in man. *Brain*, 110(2),
693 451–467. <https://doi.org/10.1093/brain/110.2.451>

694 Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of
695 single-trial EEG dynamics including independent component analysis. *Journal*

696 *of Neuroscience Methods*, 134(1), 9–21.

697 <https://doi.org/10.1016/j.jneumeth.2003.10.009>

698 Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially

699 incongruent object identification during scene viewing: Contextual expectancy

700 versus simultaneous processing. *Neuropsychologia*, 50(7), 1271–1285.

701 <https://doi.org/10.1016/j.neuropsychologia.2012.02.011>

702 Disbrow, E., Litinas, E., Recanzone, G. H., Padberg, J., & Krubitzer, L. (2003). Cortical

703 connections of the second somatosensory area and the parietal ventral area in

704 macaque monkeys. *The Journal of Comparative Neurology*, 462(4), 382–399.

705 <https://doi.org/10.1002/cne.10731>

706 Eliades, S. J., & Wang, X. (2003). Sensory-Motor Interaction in the Primate Auditory

707 Cortex During Self-Initiated Vocalizations. *Journal of Neurophysiology*, 89(4),

708 2194–2207. <https://doi.org/10.1152/jn.00627.2002>

709 Endo, N., Ito, T., Mochida, T., Ijiri, T., Watanabe, K., & Nakazawa, K. (2021). Precise

710 force controls enhance loudness discrimination of self-generated sound.

711 *Experimental Brain Research*. <https://doi.org/10.1007/s00221-020-05993-7>

712 Fuchs, M., Kastner, J., Wagner, M., Hawes, S., & Ebersole, J. S. (2002). A standardized
713 boundary element method volume conductor model. *Clinical Neurophysiology*,
714 *113*(5), 702–712. [https://doi.org/10.1016/S1388-2457\(02\)00030-5](https://doi.org/10.1016/S1388-2457(02)00030-5)

715 Han, N., Jack, B. N., Hughes, G., Elijah, R. B., & Whitford, T. J. (2021). Sensory
716 attenuation in the absence of movement: Differentiating motor action from sense
717 of agency. *Cortex*. <https://doi.org/10.1016/j.cortex.2021.04.010>

718 Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing.
719 *Nature Reviews Neuroscience*, *8*(5), 393–402. <https://doi.org/10.1038/nrn2113>

720 Ito, T., Gracco, V. L., & Ostry, D. J. (2014). Temporal factors affecting somatosensory–
721 auditory interactions in speech processing. *Frontiers in Psychology*, *5*, 1198.
722 <https://doi.org/10.3389/fpsyg.2014.01198>

723 Ito, T., Ohashi, H., & Gracco, V. L. (2021). Somatosensory contribution to audio-visual
724 speech processing. *Cortex*. <https://doi.org/10.1016/j.cortex.2021.07.013>

725 Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception.
726 *Proceedings of the National Academy of Sciences of the United States of*
727 *America*, *106*(4), 1245–1248. <https://doi.org/10.1073/pnas.0810063106>

728 Klaffehn, A. L., Baess, P., Kunde, W., & Pfister, R. (2019). Sensory attenuation prevails
729 when controlling for temporal predictability of self- and externally generated

730 tones. *Neuropsychologia*, 132, 107145.
731 <https://doi.org/10.1016/j.neuropsychologia.2019.107145>

732 Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The Cerebellum Generates
733 Motor-to-Auditory Predictions: ERP Lesion Evidence. *Journal of Cognitive*
734 *Neuroscience*, 24(3), 698–706. https://doi.org/10.1162/jocn_a_00167

735 Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction
736 of self-initiated sounds. *Cortex*, 49(9), 2449–2461.
737 <https://doi.org/10.1016/j.cortex.2012.12.012>

738 Martikainen, M. H., Kaneko, K., & Hari, R. (2005). Suppressed Responses to Self-
739 triggered Sounds in the Human Auditory Cortex. *Cerebral Cortex*, 15(3), 299–
740 302. <https://doi.org/10.1093/cercor/bhh131>

741 Mazziotta, J., Kötter, R., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods,
742 R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P.,
743 MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher,
744 N., ... Mazoyer, B. (2001). A probabilistic atlas and reference system for the
745 human brain: International Consortium for Brain Mapping (ICBM).
746 *Philosophical Transactions of the Royal Society of London. Series B: Biological*
747 *Sciences*, 356(1412), 1293–1322. <https://doi.org/10.1098/rstb.2001.0915>

- 748 Myers, J. C., Mock, J. R., & Golob, E. J. (2020). Sensorimotor Integration Can Enhance
749 Auditory Perception. *Scientific Reports*, *10*(1), 1–13.
750 <https://doi.org/10.1038/s41598-020-58447-z>
- 751 Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic
752 Response to Sound: A Review and an Analysis of the Component Structure.
753 *Psychophysiology*, *24*(4), 375–425. [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-8986.1987.tb00311.x)
754 [8986.1987.tb00311.x](https://doi.org/10.1111/j.1469-8986.1987.tb00311.x)
- 755 Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in
756 cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859.
757 <https://doi.org/10.1037/0033-2909.125.6.826>
- 758 Neszmeélyi, B., & Horváth, J. (2017). Consequences matter: Self-induced tones are used
759 as feedback to optimize tone-eliciting actions. *Psychophysiology*, *54*(6), 904–
760 915. <https://doi.org/10.1111/psyp.12845>
- 761 Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional
762 neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25.
763 <https://doi.org/10.1002/hbm.1058>

- 764 Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What Does Motor Efference
765 Copy Represent? Evidence from Speech Production. *Journal of Neuroscience*,
766 33(41), 16110–16116. <https://doi.org/10.1523/JNEUROSCI.2137-13.2013>
- 767 Ogane, R., Schwartz, J.-L., & Ito, T. (2020). Orofacial somatosensory inputs modulate
768 word segmentation in lexical decision. *Cognition*, 197, 104163.
769 <https://doi.org/10.1016/j.cognition.2019.104163>
- 770 Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain
771 dynamics. *Progress in Brain Research*, 159, 99–120.
772 [https://doi.org/10.1016/S0079-6123\(06\)59007-7](https://doi.org/10.1016/S0079-6123(06)59007-7)
- 773 Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic
774 tomography (sLORETA): Technical details. *Methods and Findings in*
775 *Experimental and Clinical Pharmacology*, 24 Suppl D, 5–12.
- 776 Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected
777 Studies of the Event-Related Potential. *International Journal of Medical*
778 *Sciences*, 2(4), 147–154.
- 779 Reznik, D., Ossmy, O., & Mukamel, R. (2015). Enhanced Auditory Evoked Activity to
780 Self-Generated Sounds Is Mediated by Primary and Supplementary Motor

781 Cortices. *Journal of Neuroscience*, 35(5), 2173–2180.
782 <https://doi.org/10.1523/JNEUROSCI.3723-14.2015>

783 Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: From sight to action.
784 *Current Opinion in Neurobiology*, 7(4), 562–567. <https://doi.org/10.1016/S0959->
785 [4388\(97\)80037-2](https://doi.org/10.1016/S0959-4388(97)80037-2)

786 Schafer, E. W. P., & Marcus, M. M. (1973). Self-Stimulation Alters Human Sensory
787 Brain Responses. *Science*, 181(4095), 175–177.
788 <https://doi.org/10.1126/science.181.4095.175>

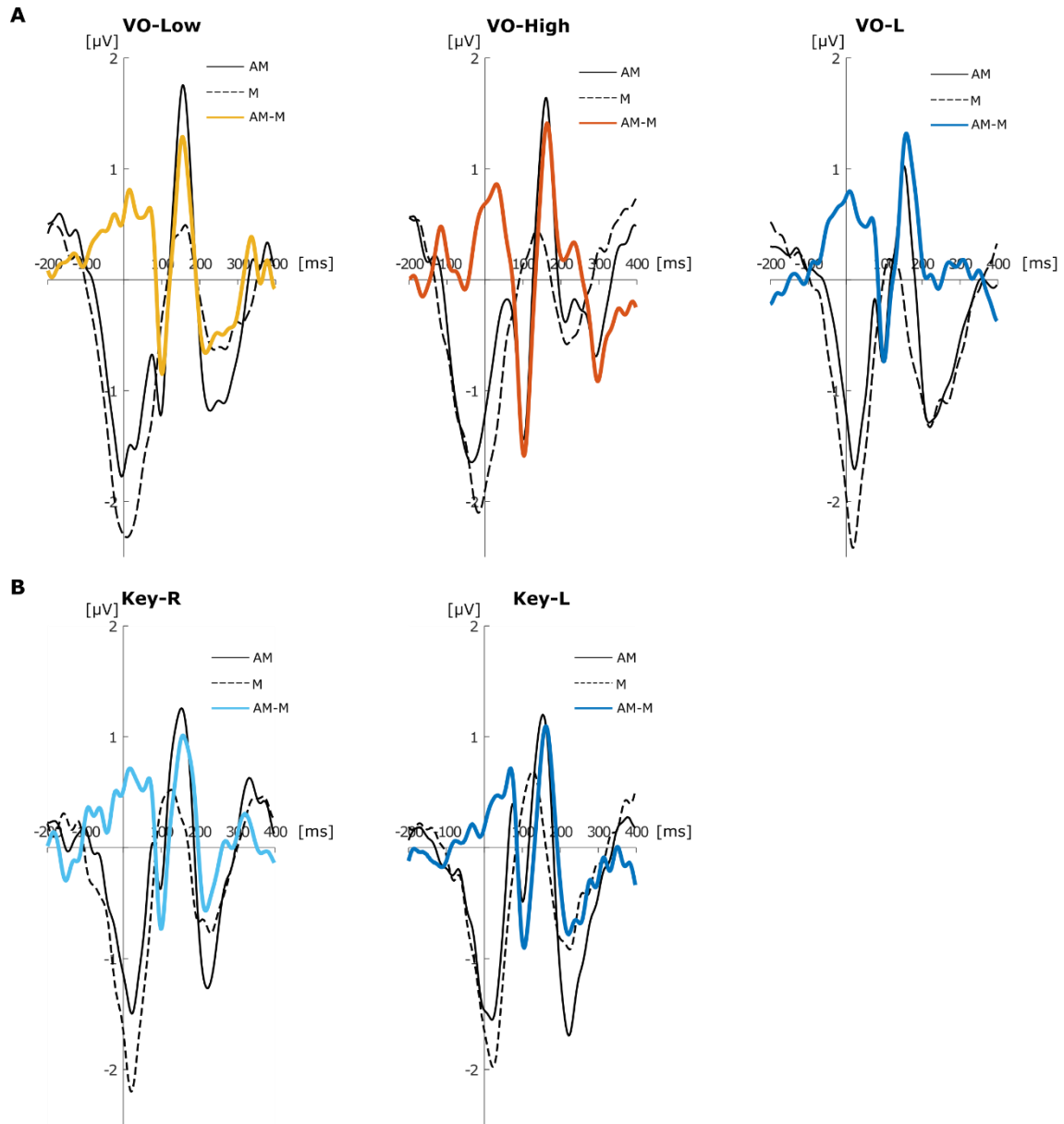
789 Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C.
790 (2001). Somatosensory Input to Auditory Association Cortex in the Macaque
791 Monkey. *Journal of Neurophysiology*, 85(3), 1322–1327.
792 <https://doi.org/10.1152/jn.2001.85.3.1322>

793 Shaffer, J. P. (1986). Modified Sequentially Rejective Multiple Test Procedures. *Journal*
794 *of the American Statistical Association*, 81(395), 826–831.
795 <https://doi.org/10.1080/01621459.1986.10478341>

796 Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by
797 visual inversion. *Journal of Comparative and Physiological Psychology*, 43(6),
798 482–489. <https://doi.org/10.1037/h0055479>

- 799 Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press.
- 800 Tachibana, R. O., Yanagida, M., & Riquimaroux, H. (2010). Novel approach for
801 understanding the neural mechanisms of auditory-motor control: Pitch
802 regulation by finger force. *Neuroscience Letters*, 482(3), 198–202.
803 <https://doi.org/10.1016/j.neulet.2010.07.032>
- 804 Tremblay, K., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the
805 auditory evoked P2 response a biomarker of learning? *Frontiers in Systems
806 Neuroscience*, 8. <https://doi.org/10.3389/fnsys.2014.00028>
- 807 Trudeau-Fisette, P., Ito, T., & Ménard, L. (2019). Auditory and Somatosensory
808 Interaction in Speech Perception in Children and Adults. *Frontiers in Human
809 Neuroscience*, 13. <https://doi.org/10.3389/fnhum.2019.00344>
- 810 Von Holst, E. (1954). Relations between the central nervous system and the peripheral
811 organs. *British Journal of Animal Behaviour*, 2, 89–94.
812 [https://doi.org/10.1016/S0950-5601\(54\)80044-X](https://doi.org/10.1016/S0950-5601(54)80044-X)
- 813 Yao, D., Qin, Y., Hu, S., Dong, L., Bringas Vega, M. L., & Valdés Sosa, P. A. (2019).
814 Which reference should we use for EEG and ERP practice?. *Brain topography*,
815 32, 530-549. <https://doi.org/10.1007/s10548-019-00707-x>

816 Yau, J., Weber, A., & Bensmaia, S. (2010). Separate Mechanisms for Audio-Tactile
817 Pitch and Loudness Interactions. *Frontiers in Psychology, 1*.
818 <https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00160>
819



821

822 **Fig. S1 Recorded auditory-motor and motor ERPs (AM ERP and M ERP) and**
823 **specific auditory responses (difference between the AM and M ERPs) in all**
824 **conditions. A) VO session (VO-Low, VO-High and VO-Left), B) Key session (Key-**
825 **Right and Key-Left).**