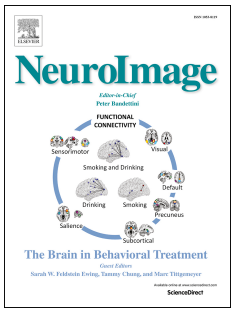
Accepted Manuscript



Inner speech is accompanied by a temporally-precise and content-specific corollary discharge

Bradley N. Jack, Mike E. Le Pelley, Nathan Han, Anthony W.F. Harris, Kevin M. Spencer, Thomas J. Whitford

PII: S1053-8119(19)30324-6

DOI: https://doi.org/10.1016/j.neuroimage.2019.04.038

Reference: YNIMG 15794

To appear in: *NeuroImage*

Received Date: 14 March 2019

Accepted Date: 11 April 2019

Please cite this article as: Jack, B.N., Le Pelley, M.E., Han, N., Harris, A.W.F., Spencer, K.M., Whitford, T.J., Inner speech is accompanied by a temporally-precise and content-specific corollary discharge, *NeuroImage* (2019), doi: https://doi.org/10.1016/j.neuroimage.2019.04.038.

Inner speech is accompanied by a temporally-precise and content-specific corollary discharge

Bradley N. Jack[1*], Mike E. Le Pelley[1], Nathan Han[1], Anthony W. F. Harris[2,3], Kevin M. Spencer[4,5], & Thomas J. Whitford[1]

1 School of Psychology, UNSW Sydney, Sydney, Australia
2 Discipline of Psychiatry, University of Sydney, Sydney, Australia
3 Brain Dynamics Centre, Westmead Institute for Medical Research, Sydney, Australia
4 Department of Psychiatry, Harvard Medical School, Boston, United States
5 Veterans Affairs Boston Healthcare System, Boston, United States

* Corresponding author: bradley.jack@unsw.edu.au

Acknowledgements: This work was supported by the Australian Research Council (DP170103094) and the National Health and Medical Research Council of Australia (APP1090507).

Abstract

When we move our articulator organs to produce overt speech, the brain generates a corollary discharge that acts to suppress the neural and perceptual responses to our speech sounds. Recent research suggests that inner speech – the silent production of words in one's mind – is also accompanied by a corollary discharge. Here, we show that this corollary discharge contains information about the temporal and physical properties of inner speech. In two experiments, participants produced an inner phoneme at a precisely-defined moment in time. An audible phoneme was presented 300 ms before, concurrently with, or 300 ms after participants produced the inner phoneme. We found that producing the inner phoneme attenuated the N1 component of the event-related potential – an index of auditory cortex processing – but only when the inner and audible phonemes occurred concurrently and matched on content. If the audible phoneme was presented before or after the production of the inner phoneme, or if the inner phoneme did not match the content of the audible phoneme, there was no attenuation of the N1. These results suggest that inner speech is accompanied by a temporally-precise and content-specific corollary discharge. We conclude that these results support the notion of a functional equivalence between the neural processes that underlie the production of inner and overt speech, and provide empirical support for the influential hypothesis that inner speech is a special form of overt speech.

Keywords: Inner speech, internal forward model, N1, event-related potentials (ERPs).

Introduction

As you read this text, you can probably hear your inner voice narrating the words. Inner speech – the silent production of words in one's mind (Alderson-Day & Fernyhough, 2015; Perrone-Bertolotti et al., 2014; Zivin, 1979) – is a core aspect of our mental lives; it is linked to a wide-range of psychological functions, including reading, writing, planning, memory, self-motivation, and problem-solving (Alderson-Day et al., 2018; Morin et al., 2011, 2018; Sokolov et al., 1972). Despite its ubiquity, relatively little is known about the neural processes that underlie the production of inner speech. One influential hypothesis states that inner speech is a special form of overt speech (Feinberg, 1978; Frith, 1987; Jones & Fernyhough, 2007). Evidence for this comes from the observation that the brain regions involved in producing inner speech are similar to those involved in producing overt speech, including auditory, language, and supplementary motor areas (Aleman et al., 2005; McGuire et al., 1996; Palmer et al., 2001; Shergill et al., 2001; Shuster & Lemieux, 2005; Zatorre et al., 1996). According to the internal forward model of overt speech (Miall & Wolpert, 1996), when we move our articulator organs to speak, an *efference copy* is issued in parallel (Von Holst & Mittelstaedt, 1950). This efference copy forms the basis of a neural prediction – a *corollary discharge* (Sperry, 1950) – regarding the temporal and physical properties of our speech sounds, which is used to suppress the neural and perceptual responses to those sounds (Crapse & Sommer, 2008; Straka et al., 2018). If inner speech is, in fact, a special form of overt speech, then it should also be accompanied by a temporally-precise and content-specific corollary discharge. The present study investigated this issue.

There is a growing body of research suggesting that inner speech is accompanied by a corollary discharge (Ford & Mathalon, 2004; Scott, 2013; Tian & Poeppel, 2010, 2012, 2013, 2015; Tian et al., 2016, 2018; Whitford et al., 2017; Ylinen et al., 2015). Of particular relevance to the present study is an experiment conducted by Whitford et al. (2017), who introduced a procedure in which participants viewed a ticker-tape-style cue which provided them with precise knowledge about when they would hear an audible phoneme. In the *listen condition* of their experiment, participants were instructed to passively listen to the audible phoneme; in the *inner speech condition*, participants were instructed to produce an inner phoneme at the precise moment they heard the audible phoneme. On a random half of the trials in the inner speech condition, the inner and audible phonemes matched on content – this was called the *match condition*; on the other half of the trials, the inner and audible phonemes did not match on content – this was called the *mismatch condition*. Whitford et al. (2017) found that producing the inner phoneme attenuated the N1 component of the event-related potential (ERP) – an index of auditory cortex processing (Näätänen & Picton, 1987; Woods, 1995) – compared to passive listening, but only when the inner and audible phonemes matched on content. If the inner phoneme did not match the content of the audible phoneme, there was no attenuation of the N1. These results suggest that inner speech, similar to overt speech (Behroozmand et al., 2009;

Behroozmand & Larson, 2011; Eliades & Wang, 2008; Heinks-Maldonado et al., 2005; Houde et al., 2002; Liu et al., 2011; Sitek et al., 2013), is accompanied by a content-specific corollary discharge, in that it contains information about the physical properties of inner speech.

However, when we move our articulator organs to speak, the accompanying corollary discharge is not only content-specific, but also temporally-precise, in that it contains information about the temporal properties of overt speech. Evidence for this comes from studies showing that N1-attenuation can be reduced or abolished by imposing a temporal delay between articulator movement and auditory feedback (Behroozmand et al., 2010, 2016; Chen et al., 2012; see also Blakemore et al., 1999; Elijah et al., 2016; Oestreich et al., 2016; Whitford et al., 2011). In the present study, we investigated whether inner speech, like overt speech, is accompanied by a temporally-precise and content-specific corollary discharge. To accomplish this, we used the same ticker-tape-style cue introduced by Whitford et al. (2017) to control the time at which participants produced the inner phoneme, and we presented the audible phoneme 300 ms before, concurrently with, or 300 ms after participants produced the inner phoneme – we call these the *before*, *precise*, and *after conditions*, respectively. In Experiment 1, we compared the N1 elicited by the audible phoneme during passive listening and the production of inner speech across the different time delays; in Experiment 2, we compared the N1 elicited by an audible phoneme that either matched or mismatched the inner phoneme across the different time delays. Assuming that inner speech is accompanied by a temporally-precise and content-specific corollary discharge, we hypothesize larger N1-attenuation effects when the timing and content of the inner phoneme matches the audible phoneme compared to when it does not.

Experiment 1

Method

Participants. Forty-two students from UNSW Sydney participated in our study for course credit. All participants gave written informed consent prior to the experiment and reported having normal hearing in both ears. Data from three participants were excluded from the analyses due to excessive artefacts in the electroencephalogram **(**EEG) recording (> 75% of epochs meeting the rejection criteria; see ERP processing and ERP analysis). Mean age of the remaining participants, 20 of whom were female and 38 of whom were right-handed, was 20 (*SD* = 3) years. The experiment was approved by UNSW Sydney's Human Research Ethics Advisory Panel and was conducted in accordance with the ethical standards laid down in the Declaration of Helsinki (World Medical Association, 2004).

Apparatus, stimuli, and procedure. Participants sat in a quiet, dimly-lit room, approximately 60 cm in front of a computer monitor (BenQ XL2420T) and wore headphones (AKG K77). Stimulus presentation was controlled by specially written Matlab scripts using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Participants watched an animation, on every trial: it

began with a green horizontal line in the centre of the screen – the ticker-tape – a red vertical line in the centre of the screen – the fixation line – and a green vertical line on the right-hand side of the screen – the target line (see Figure 1a). Participants were instructed to look at the fixation line (which remained stationary) for the duration of the trial. After a 1 s delay, the target line began to move leftwards across the screen at a speed of 6.5 °/s, such that after 4 s the target line overlapped the fixation line and subsequently continued to move across the ticker-tape for an additional 1 s (see Figure 1b–f). After each trial, participants rated their subjective performance on that trial with a 5-point Likert scale, with scores ranging from 1, meaning "not at all successful", to 5, meaning "completely successful". We used these ratings to identify and classify trials in which participants successfully performed the task.

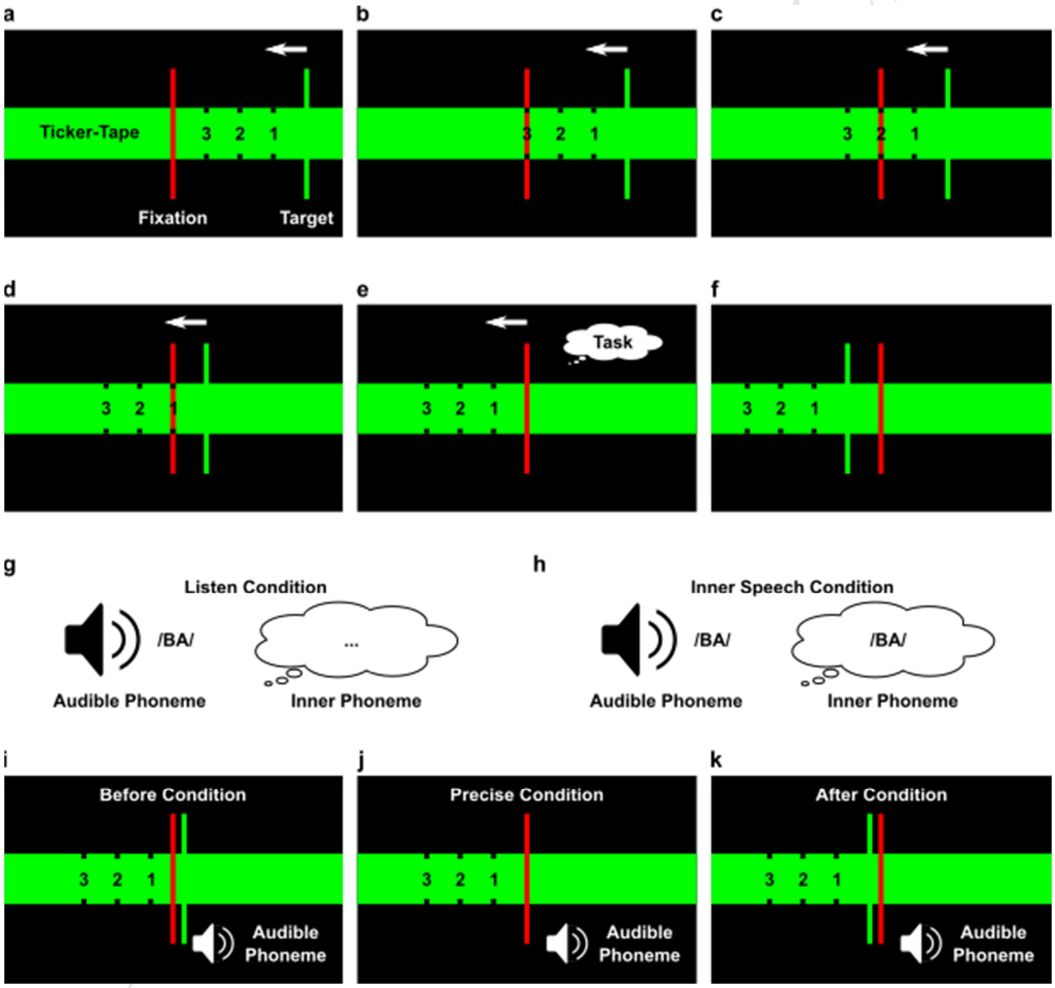


Figure 1. Procedure for Experiment 1. (a–f) Participants were instructed to look at the fixation line (which remained stationary) for the duration of the trial. After a short delay, the target line began to move leftwards across the screen such that after 4 s the target line overlapped the fixation line and subsequently continued to move across the ticker-tape. In the *listen condition*, participants were instructed to passively listen to a recording of the phoneme /ba/. (h) In the *inner speech condition*, participants were instructed to silently produce the phoneme /ba/ in their minds at the precise moment the fixation and target lines overlapped (as shown in e). (i) On a random one-third of trials for

both conditions, the audible phoneme was presented 300 ms before the fixation and target lines overlapped – the *before condition*; (j) on a different one-third of trials, the audible phoneme was presented at the precise moment the fixation and target lines overlapped – the *precise condition*; (k) on the remaining one-third of trials, the audible phoneme was presented 300 ms after the fixation and target lines overlapped – the *after condition*.

The experiment consisted of 20 blocks of trials, with each block containing 18 trials. On half of the blocks, participants performed the *listen condition*: they were instructed to passively listen to a recording of the audible phoneme /ba/ (see Figure 1g), which was produced by a male speaker, was about 200 ms long, and was about 70 dB SPL. On the other half of the blocks, participants performed the *inner speech condition*: they listened to a recording of the audible phoneme /ba/, and they were instructed to silently produce the phoneme /ba/ in their minds at the precise moment the fixation and target lines overlapped (see Figure 1h). The order of blocks alternated between the listen and inner speech conditions, and the starting block was counterbalanced over participants.

On a random one-third of the trials in each block, the audible phoneme was presented 300 ms before the fixation and target lines overlapped – the *before condition* (see Figure 1i); on a different one-third of the trials, the audible phoneme was presented at the precise moment the fixation and target lines overlapped – the *precise condition* (see Figure 1j); on the remaining one-third of the trials, the audible phoneme was presented 300 ms after the fixation and target lines overlapped – the *after condition* (see Figure 1k). The order of the before, precise, and after conditions was random and different for each block, as well as different for each participant.

EEG acquisition. We recorded the EEG with a BioSemi ActiveTwo system using 64 Ag/AgCl active electrodes placed according to the extended 10-20 system (FP1, FPz, FP2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P9, P7, P5, P3, P1, Pz, P2, P4, P6, P8, P10, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, Iz). We also recorded the vertical electrooculogram (EOG) by placing an electrode above (we used FP1) and below the left eye, and the horizontal EOG by placing an electrode on the outer canthus of each eye. We also placed an electrode on the tip of the nose. The sampling rate of the EEG was 2,048 Hz.

ERP processing and ERP analysis. We re-referenced the data to the electrode on the tip of the nose, and we filtered the data using a half-amplitude 0.5 to 30 Hz phase-shift free Butterworth filter (48 dB/Oct slope), as well as a 50 Hz Notch filter. We extracted the epochs from -100 to 400 ms relative to audible phoneme onset, we corrected the epochs for eye-blink and movement artefacts using the technique described in Gratton et al. (1983) and Miller et al. (1988), and we excluded all epochs with signals exceeding peak-to-peak amplitudes of 200 µV at any EEG channel. We also excluded any epochs in which participants subsequently rated their performance on the trial as less than or equal to 3 out of

5. We baseline-corrected all epochs to their mean voltage from –100 to 0 ms, and we computed an ERP for each condition. On average, ERPs were computed from 43 (*SD* = 16) *listen-before*, 52 (*SD* = 10) *listen-precise*, 37 (*SD* = 12) *listen-after*, 34 (*SD* = 18) *inner speech-before*, 48 (*SD* = 15) *inner speech-precise*, and 33 (*SD* = 14) *inner speech-after* epochs. We analysed the mean amplitude of the N1 averaged over Fz, FCz, and Cz electrodes in the time-window of 80 to 120 ms with repeated-measure ANOVA using the factors *task* (listen, inner speech) and *time* (before, precise, after). We chose these electrodes to be consistent with Whitford et al. (2017) and the literature on N1-attenuation to overt speech (Behroozmand et al., 2009, 2010, 2016; Behroozmand & Larson, 2011; Chen et al., 2012; Eliades & Wang, 2008; Heinks-Maldonado et al., 2005; Houde et al., 2002; Liu et al., 2011; Sitek et al., 2013), and we selected this time-window using the collapsed localiser technique (Luck & Gaspelin, 2017).

Results

Behavioural results. Participants rated their subjective performance after each trial with a 5-point Likert scale, with scores ranging from 1, meaning "not at all successful", to 5, meaning "completely successful". Participants' mean ratings were 4.12 (*SD* = 0.69) in the listen-before condition, 4.60 (*SD* = 0.46) in the listen-precise condition, 4.30 (*SD* = 0.56) in the listen-after condition, 3.52 (*SD* = 0.86) in the inner speech-before condition, 4.37 (*SD* = 0.74) in the inner speech-precise condition, and 4.14 (*SD* = 0.80) in the inner speech-after condition. These results show that participants performed the task as instructed.

ERP results. Figure 2a shows the ERPs, Figure 2b shows the mean amplitudes for the N1 time-window, and Figure 2c shows the voltage maps for the N1 time-window. Repeated-measures ANOVA found a significant interaction between task and time, $F(2, 76) = 3.84$, $p = .026$, $\eta_p^2 = .09$. There was also a significant main effect of time, $F(2, 76) = 3.94$, $p = .024$, $\eta_p^2 = .09$; however, the main effect of task was not significant, $F(1, 38) = 0.21$, $p = .649$, $\eta_p^2 < .01$. Post-hoc *t*-tests found that N1-amplitude was significantly smaller for the inner speech-precise condition than for the listen-precise condition, $t(38) = 2.64$, $p = .012$, $d = 0.42$. However, the difference between the inner speech-before and listen-before conditions was not significant, $t(38) = 0.88$, $p = .383$, $d = 0.14$, nor was the difference between the inner speech-after and listen-after conditions, $t(38) = 0.69$, $p = .496$, $d = 0.11$. Moreover, N1-amplitude was significantly smaller for the inner speech-precise condition than for the inner speech-before condition, $t(38) = 3.64$, $p = .001$, $d = 0.58$, and for the inner speech-after condition, $t(38) = 3.03$, $p = .004$, $d = 0.49$. There were no other significant differences. These results show that producing the inner phoneme attenuated the N1 compared to passive listening, but only when the inner and audible phonemes occurred concurrently. If the audible phoneme was presented before or after the production of the inner phoneme, there was no attenuation of the N1. This pattern of results is consistent with the idea that inner speech, similar to overt speech, is accompanied by a temporally-precise corollary discharge.

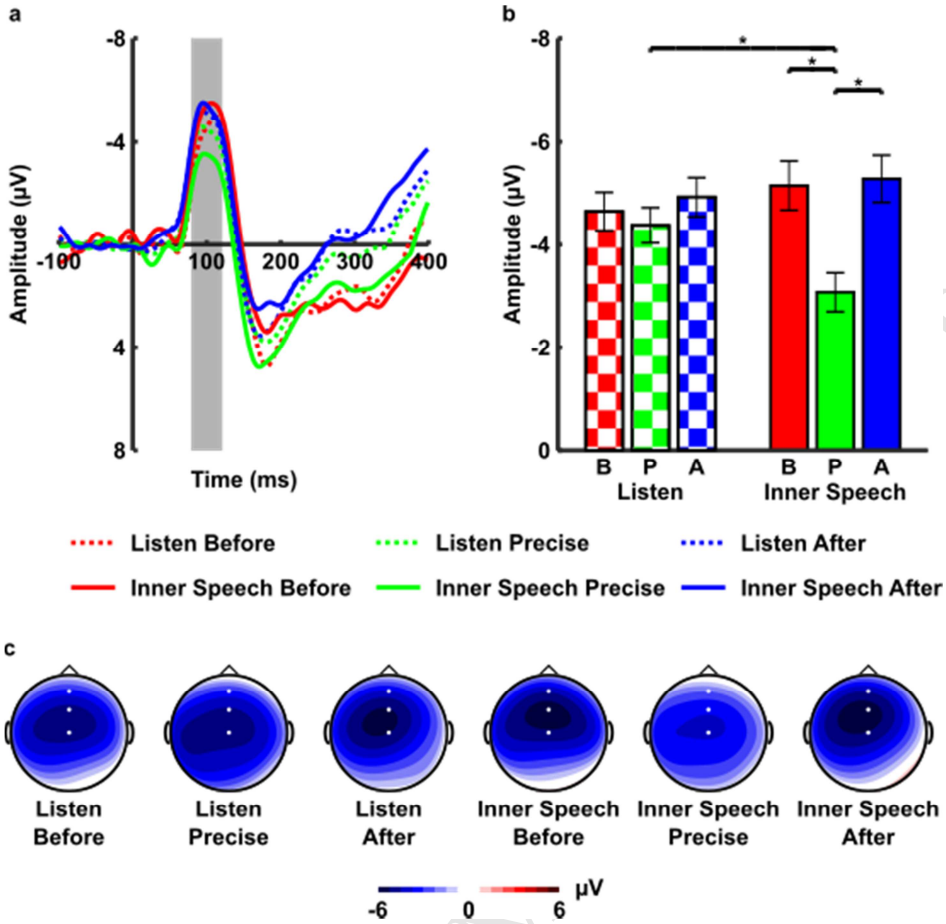


Figure 2. Results for Experiment 1. (a) The graph shows the grand-averaged ERPs for each condition averaged over Fz, FCz, and Cz electrodes, showing time (ms) on the *x*-axis, with 0 indicating the onset of the auditory phoneme, and voltage (μV) on the *y*-axis, with negative voltages plotted upwards. The grey bar shows the N1 time-window (80–120 ms), which was selected using the collapsed localiser technique (Luck & Gaspelin, 2017). (b) The bar graph shows the mean amplitudes for the N1 time-window for the listen and inner speech conditions across the different time delays: before (B), precise (P), and late (L). Error bars show the standard error of the mean (SEM). (c) The voltage maps show the distribution of voltages over the scalp during the N1 time-window.

Our primary focus is N1-amplitude; however, we also conducted supplementary analyses on the peak latency of the N1 and the mean amplitudes of the P2 and P3. To see the results of these analyses, see Appendix A.

Experiment 2

Method

Participants. Sixty-one students participated in our study for course credit. Data from six participants were excluded from the analyses due to excessive artefacts in the EEG recording (see ERP processing and ERP analysis). Mean age of the remaining participants, 42 of whom were female and 52 of whom were right-handed, was 20 (*SD* = 3) years.

Apparatus, stimuli, and procedure. The apparatus, stimuli, and animation were identical to Experiment 1. The experiment consisted of 20 blocks of trials,

with each block containing 18 trials. On half of the blocks, participants performed the *inner speech /ba/ condition*: they were instructed to silently produce the phoneme /ba/ in their minds at the precise moment the fixation and target lines overlapped; on the other half of the blocks, participants performed the *inner speech /bi/ condition*: they were instructed to silently produce the phoneme /bi/ in their minds at the precise moment the fixation and target lines overlapped. The order of the blocks alternated between the inner speech /ba/ and inner speech /bi/ conditions, and the starting block was counterbalanced over participants.

On a random half of the trials in each block, the inner and audible phonemes matched on content; that is, participants produced the phoneme /ba/ or /bi/ and listened to a recording of the phoneme /ba/ or /bi/, respectively – the *match condition* (see Figure 3a–b). On the other half of trials, the inner and audible phonemes did not match on content; that is, participants produced the phoneme /ba/ or /bi/ and listened to a recording of the phoneme /bi/ or /ba/, respectively – the *mismatch condition* (see Figure 3c–d). Similar to Experiment 1, on a random one-third of trials for the match and mismatch conditions, the audible phoneme was presented 300 ms before the fixation and target lines overlapped – the *before condition*; on a different one-third of the trials, the audible phoneme was presented at the precise moment the fixation and target lines overlapped – the *precise condition*; on the remaining one-third of the trials, the audible phoneme was presented 300 ms after the fixation and target lines overlapped – the *after condition*. The order of the trials was random and different for each block, as well as different for each participant.

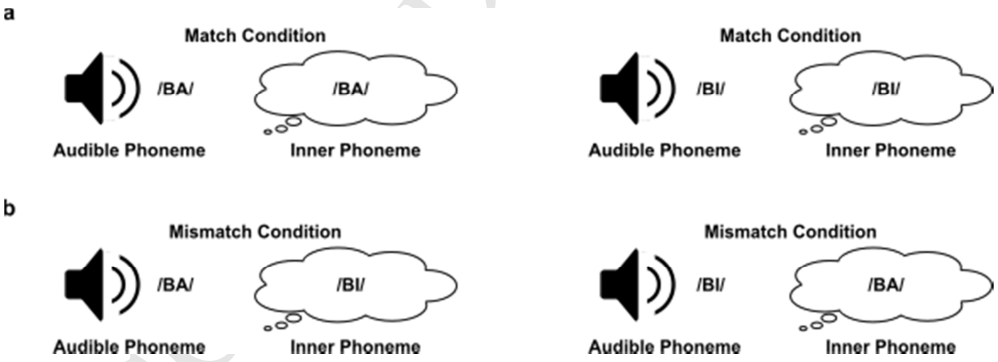


Figure 3. Procedure for Experiment 2. (a–b) On half of the blocks, participants were instructed to silently produce the phoneme /ba/ in their minds at the precise moment the fixation and target lines overlapped; on the other half of the blocks, participants were instructed to silently produce the phoneme /bi/ in their minds at the precise moment the fixation and target lines overlapped. On half of the trials in each block, the inner and audible phonemes matched on content – the *match condition*; (c–d) on the other half of trials, the inner and audible phonemes did not match on content – the *mismatch condition*. Similar to Experiment 1, on a random one-third of trials for both conditions, the audible phoneme was presented 300 ms before the fixation and target lines overlapped – the *before condition*; on a different one-third of trials, the audible phoneme was presented at the precise moment the fixation and target lines overlapped – the *precise condition*; on the remaining one-third of

trials, the audible phoneme was presented 300 ms after the fixation and target lines overlapped – the *after condition*.

EEG acquisition. The EEG acquisition was identical to Experiment 1.

ERP processing and ERP analysis. The ERP processing and ERP analysis were identical to Experiment 1. On average, ERPs were computed from 39 (*SD* = 18) *match-before*, 50 (*SD* = 10) *match-precise*, 31 (*SD* = 14) *match-after*, 31 (*SD* = 16) *mismatch-before*, 42 (*SD* = 15) *mismatch-precise*, and 31 (*SD* = 13) *mismatch-after* epochs. Similar to Experiment 1, we analysed the mean amplitude of the N1 averaged over Fz, FCz, and Cz electrodes in the time-window of 80 to 120 ms with repeated-measure ANOVA using the factors *task* (listen, inner speech) and *time* (before, precise, after).

Results

Behavioural results. Participants' mean ratings were 3.92 (*SD* = 0.80) in the match-before condition, 4.54 (*SD* = 0.43) in the match-precise condition, 4.38 (*SD* = 0.52) in the match-after condition, 3.44 (*SD* = 0.80) in the mismatch-before condition, 4.00 (*SD* = 0.76) in the mismatch-precise condition, and 4.03 (*SD* = 0.75) in the mismatch-after condition. Again, these results show that participants performed the task as instructed.

ERP results. Figure 4a shows the ERPs, Figure 4b shows the mean amplitudes for the N1 time-window, and Figure 4c shows the voltage maps for the N1 time-window. Repeated-measures ANOVA found a significant interaction between task and time, $F(2, 108) = 3.25$, $p = .043$, $\eta_p^2 = .06$. There was also a significant main effect of time, $F(2, 108) = 6.84$, $p = .002$, $\eta_p^2 = .11$; however, the main effect of task was not significant, $F(1, 54) = 0.03$, $p = .856$, $\eta_p^2 < .01$. Post-hoc *t*-tests found that N1-amplitude was significantly smaller for the match-precise condition than for the mismatch-precise condition, $t(54) = 2.38$, $p = .021$, $d = 0.32$. However, the difference between the match-before and mismatch-before conditions was not significant, $t(54) = 1.43$, $p = .160$, $d = 0.19$, nor was the difference between the match-after and mismatch-after conditions, $t(54) = 0.62$, $p = .536$, $d = 0.08$. Moreover, N1-amplitude was significantly smaller for the match-precise condition than for the match-before condition, $t(54) = 5.63$, $p < .001$, $d = 0.76$, and for the match-after condition, $t(54) = 2.08$, $p = .043$, $d = 0.28$, as well as significantly smaller for the match-after condition than for the match-before condition, $t(54) = 2.40$, $p = .020$, $d = 0.32$. There were no other significant differences. These results show that producing an inner phoneme that matched the audible phoneme attenuated the N1 compared to when the inner and audible phonemes did not match, but only when the inner and audible phonemes occurred concurrently. If the audible phoneme was presented before or after the production of the inner phoneme, there was no attenuation of the N1. This pattern of results is consistent with the idea that inner speech, similar to overt speech, is accompanied by a temporally-precise and content-specific corollary discharge.

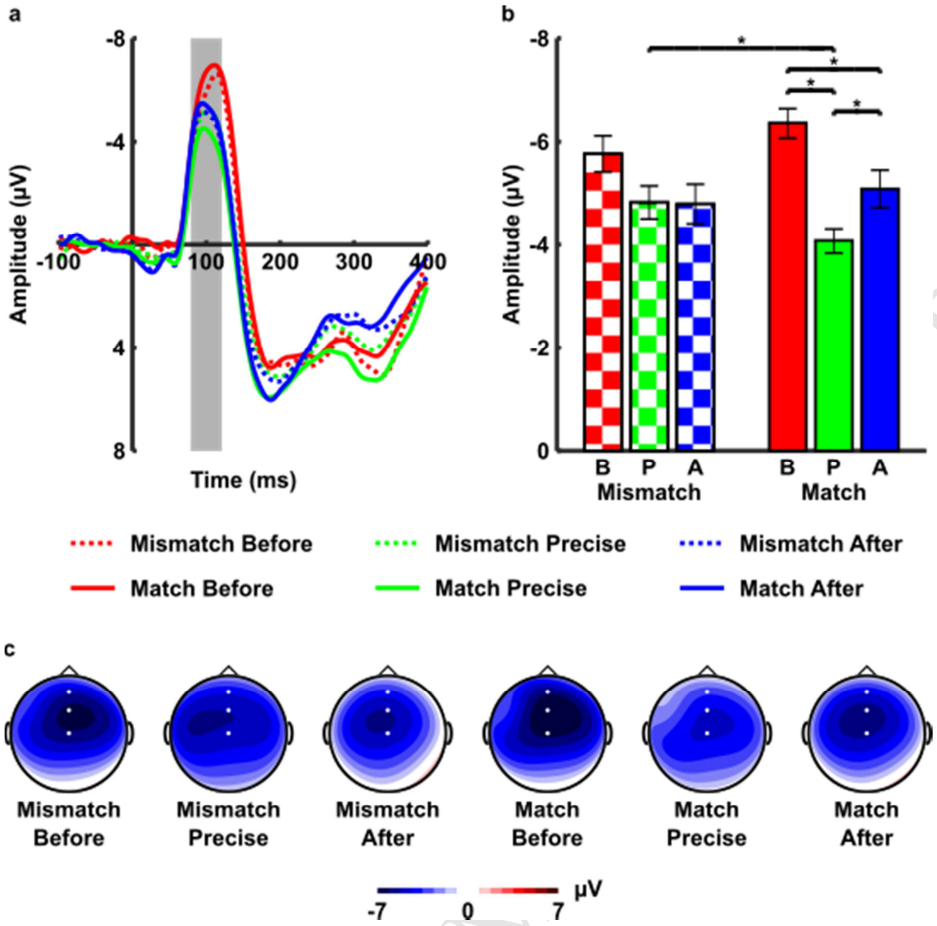


Figure 4. Results for Experiment 2. (a) The graph shows the grand-averaged ERPs for each condition averaged over Fz, FCz, and Cz electrodes, showing time (ms) on the *x*-axis, with 0 indicating the onset of the auditory phoneme, and voltage (μV) on the *y*-axis, with negative voltages plotted upwards. The grey bar shows the N1 time-window (80–120 ms), which we used to be consistent with Experiment 1. (b) The bar graph shows the mean amplitudes for the N1 time-window for the match and mismatch conditions across the different time delays: before (B), precise (P), and late (L). Error bars show the SEM. (c) The voltage maps show the distribution of voltages over the scalp during the N1 time-window.

Similar to Experiment 1, we also conducted supplementary analyses on the peak latency of the N1 and the mean amplitudes of the P2 and P3. To see the results of these analyses, see Appendix B.

Discussion

We set out to determine the properties of the corollary discharge associated with inner speech: specifically, whether it contains information about the temporal and physical properties of inner speech. In two experiments, participants produced an inner phoneme at a precisely-defined moment in time, and an audible phoneme was presented 300 ms before, concurrently with, or 300 ms after participants produced the inner phoneme. The results of the present study were unequivocal: we found that producing the inner phoneme attenuated the N1, but only when the inner and audible phonemes occurred concurrently and matched on content. If the audible phoneme was presented before or after the production of the inner

phoneme, or if the inner phoneme did not match the content of the audible phoneme, there was no attenuation of the N1. These results suggest that inner speech, similar to overt speech (Behroozmand et al., 2009, 2010, 2016; Behroozmand & Larson, 2011; Chen et al., 2012; Eliades & Wang, 2008; Heinks-Maldonado et al., 2005; Houde et al., 2002; Liu et al., 2011; Sitek et al., 2013), is accompanied by a corollary discharge that is both temporally-precise and content-specific. We conclude that these results support the notion of a functional equivalence between the neural processes that underlie the production of inner and overt speech, and provide empirical support for the influential, yet relatively untested, hypothesis that inner speech is a special form of overt speech (Feinberg, 1978; Frith, 1987; Jones & Fernyhough, 2007).

To the best of our knowledge, only one other study has attempted to investigate the temporal precision of inner speech. Tian and Poeppel (2015) asked their participants to press a button at the precise moment they produced an inner phoneme. An audible phoneme that matched the content of the inner phoneme was presented concurrently with, 100, 200, or 500 ms after the button-press. Tian and Poeppel (2015) found attenuation of the M1 (the magnetoencephalogram equivalent of the N1; Virtanen et al., 1998) when the inner and audible phonemes occurred concurrently and when the delay between them was 100 ms, but not when the delay was 200 or 500 ms. These results are consistent with ours in that we found N1-attenuation when the inner and audible phonemes occurred concurrently, but not when the delay was 300 ms. However, the present study represents an important departure from Tian and Poeppel (2015). Specifically, their participants pressed a button to signal the production of the inner phoneme. This aspect of their procedure is potentially problematic, because finger movements (such as those involved in pressing a button) are known to attenuate the M1 and N1 of the auditory-evoked potential (Aliu et al., 2009; Bäß et al., 2008; Blakemore et al., 1999; Elijah et al., 2016; Knolle et al., 2013; Mifsud et al., 2016; Oestreich et al., 2016; SanMiguel et al., 2013; Timm et al., 2013; Whitford et al., 2011), which makes it difficult to determine whether the M1 reductions observed by Tian and Poeppel (2015) were caused by the inner speech, the button-press, or some combination of the two. Furthermore, finger movements produce a motor-evoked potential. This makes it difficult to determine whether the M1 reductions reflected suppression of the auditory-evoked potential elicited by the audible phoneme, the motor-evoked potential elicited by the button-press, or some combination of the two. In contrast, our procedure did not require participants to press a button to signal the production of the inner phoneme; instead, they watched an animation and produced the inner phoneme at a precisely-defined moment in time. By eliminating the need for a button-press, the present study provides the most convincing evidence yet that inner speech is accompanied by a temporally-precise corollary discharge.

The results of the present study suggest that the corollary discharge associated with inner speech does not result in broad, blanket suppression of all auditory input over an extended period; rather, it suppresses the input that matches

the content of inner speech at the precise moment that it is "spoken". This pattern of results has previously been reported in studies of overt speech (Behroozmand et al., 2009, 2010, 2016; Behroozmand & Larson, 2011; Chen et al., 2012; Eliades & Wang, 2008; Heinks-Maldonado et al., 2005; Houde et al., 2002; Liu et al., 2011; Sitek et al., 2013), and is typically interpreted in the context of the internal forward model (Miall & Wolpert, 1996). According to this framework, the brain uses a corollary discharge to predict the sensory consequences of the movement of our articulator organs and to suppress the auditory input consistent with this prediction (Crapse & Sommer, 2008; Straka et al., 2018). The results of the present study suggest that inner speech exerts a similar effect on auditory processing, indicating a functional equivalence between the corollary discharges associated with inner and overt speech, even though inner speech does not produce an audible sound. In this sense, our results demonstrate a case in which the brain's prediction goes too far, generating an expectation of a sensory event that does not occur. This prompts the following question: *why* is inner speech accompanied by a corollary discharge? We suspect that the most likely explanation is that inner speech evolved from overt speech, and thus continued to use many of the same underlying neural processes, including corollary discharges (Alderson-Day & Fernyhough, 2015; Jones & Fernyhough, 2007); however, we concede that this possibility is speculation.

The results of the present study also support the influential hypothesis that inner speech is a special form of overt speech (Feinberg, 1978; Frith, 1987; Jones & Fernyhough, 2007), in that both yield similar effects on auditory processing. This lends support to the intriguing suggestion that the brain does not make a conceptual distinction between thoughts and actions, at least in the context of speech. But does this extend to situations involving non-speech actions? For example, does thinking about making a hand or finger movement result in N1-attenuation to a consequential sound, similar to what has been observed in response to actual hand or finger movements? Recent research from Kilteni et al. (2018) suggests that content-specific corollary discharges may accompany imagined hand and finger movements; however, more research is needed. Finally, the present study has important implications beyond our understanding of the neurobiology of thoughts. For instance, dysfunctions of inner speech (Feinberg, 1978; Frith, 1987) – and specifically, dysfunctions in the *timing* of inner speech (Whitford et al., 2011, 2012) – have been argued to underlie certain classes of auditory-verbal hallucinations, such as audible thoughts (*Gedankenlautwerden*), which are highly characteristic of schizophrenia (Fletcher & Frith, 2009; Mellor, 1970). Our procedure allows us to quantify the timing of inner speech by measuring its effect on auditory processing. As such, it unlocks the possibility of directly testing the long-held, but hitherto untested, hypothesis regarding the critical role of inner speech dysfunction in auditory-verbal hallucinations. Our procedure may also be useful for the ongoing development of brain-computer interfaces aimed at deciphering inner speech for people who are unable to produce overt speech (Lebedev & Nicolelis, 2006).

In summary, we investigated whether inner speech is accompanied by a temporally-precise and content-specific corollary discharge. In two experiments, we found electrophysiological evidence in support of this possibility. Specifically, we found that producing the inner phoneme attenuated the N1, but only when the inner and audible phonemes occurred concurrently and matched on content. If the audible phoneme was presented before or after the production of the inner phoneme, or if the inner phoneme did not match the content of the audible phoneme, there was no attenuation of the N1. These results replicate and extend upon Whitford et al. (2017) and Tian and Poeppel (2015), and suggest that inner speech, similar to overt speech (Behroozmand et al., 2009, 2010, 2016; Behroozmand & Larson, 2011; Chen et al., 2012; Eliades & Wang, 2008; Heinks-Maldonado et al., 2005; Houde et al., 2002; Liu et al., 2011; Sitek et al., 2013), is accompanied by a corollary discharge that is both temporally-precise and content-specific. We conclude that these results support the notion of a functional equivalence between the neural processes – namely, efference copies and corollary discharges – that underlie the production of inner and overt speech, and provide empirical support for one of the most influential hypotheses in cognitive neuroscience: that inner speech is a special form of overt speech (Feinberg, 1978; Frith, 1987; Jones & Fernyhough, 2007).

Appendix A: Supplementary analyses for Experiment 1

We conducted three supplementary analyses that were not directly related to our hypotheses. First, we analysed the peak latency of the N1 by identifying the most negative voltage averaged over Fz, FCz, and Cz electrodes in the time-window of 50 to 150 ms for every condition and participant. Repeated-measures ANOVA found that the main effect of task was not significant, $F(1, 38) = 0.19$, $p = .665$, $\eta_p^2 < .01$, that the main effect of time was not significant, $F(2, 76) = 0.69$, $p = .504$, $\eta_p^2 = .02$, and that the interaction between task and time was not significant, $F(2, 76) = 0.54$, $p = .586$, $\eta_p^2 = .01$.

We then analysed the mean amplitude of the P2 (Crowley & Colrain, 2004) averaged over FCz, Cz, and CPz electrodes in the time-window of 160 to 200 ms. We chose these electrodes to be consistent with Whitford et al. (2017) and we selected this time-window using the collapsed localiser technique (Luck & Gaspelin, 2017). Figure S1a shows the ERPs, Figure S1b shows the mean amplitudes for the P2 time-window, and Figure S1c shows the voltage maps for the P2 time-window. Repeated-measures ANOVA found a significant interaction between task and time, $F(2, 76) = 3.29$, $p = .043$, $\eta_p^2 = .08$; however, the main effect of task was not significant, $F(1, 38) = 0.15$, $p = .698$, $\eta_p^2 < .01$, and the main effect of time was not significant, $F(2, 76) = 0.74$, $p = .480$, $\eta_p^2 = .02$. Post-hoc *t*-tests found that P2-amplitude was significantly larger for the inner speech-precise condition than for the inner speech-after condition, $t(38) = 2.13$, $p = .040$, $d = 0.34$. There were no other significant differences.

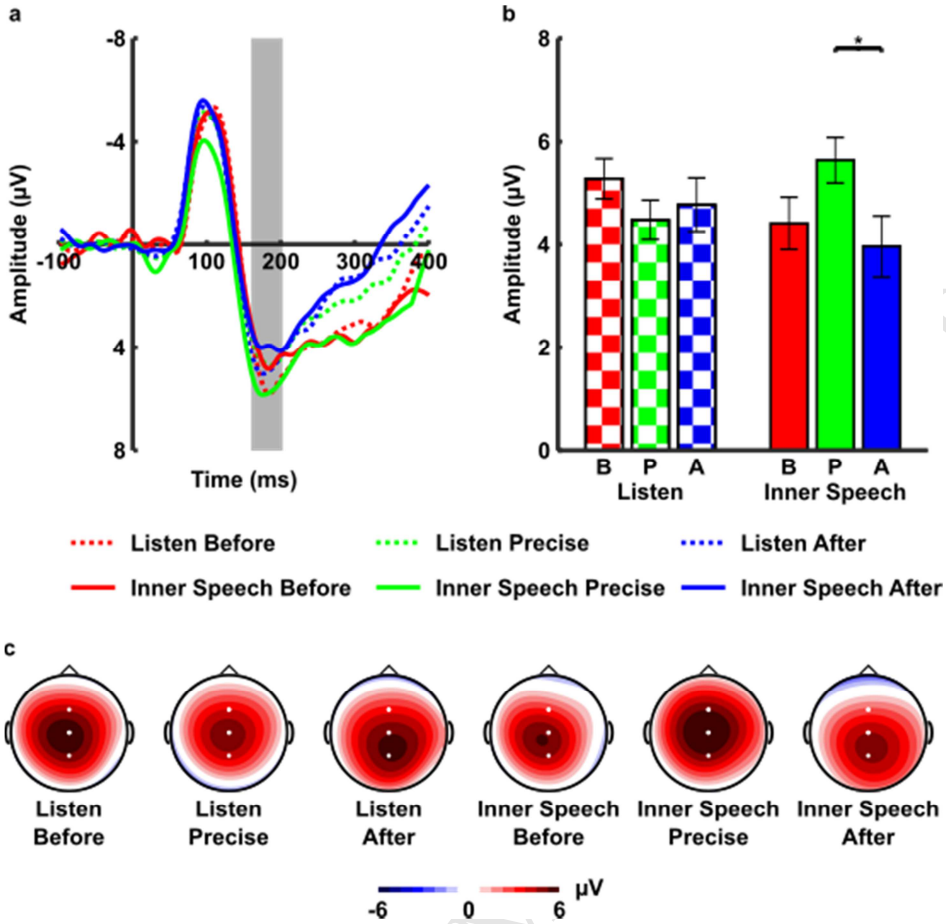


Figure S1. Analysis of the P2. (a) The graph shows the grand-averaged ERPs for each condition averaged over FCz, Cz, and CPz electrodes, showing time (ms) on the *x*-axis, with 0 indicating the onset of the auditory phoneme, and voltage (μV) on the *y*-axis, with negative voltages plotted upwards. The grey bar shows the P2 time-window (160–200 ms), which was selected using the collapsed localiser technique (Luck & Gaspelin, 2017). (b) The bar graph shows the mean amplitudes for the P2 time-window for the listen and inner speech conditions across the different time delays: before (B), precise (P), and late (L). Error bars show the SEM. (c) The voltage maps show the distribution of voltages over the scalp during the P2 time-window.

Finally, we analysed the mean amplitude of the P3 (Polich, 2007) averaged over Cz, CPz, and Pz electrodes in the time-window of 250 to 350 ms. We chose these electrodes to be consistent with Whitford et al. (2017) and we selected this time-window after visual inspection of the ERPs and voltage maps, because there was no discernible P3-peak in the ERPs. Figure S2a shows the ERPs, Figure S2b shows the mean amplitudes for the P3 time-window, and Figure S2c shows the voltage maps for the P3 time-window. Repeated-measures ANOVA found that the main effect of task was not significant, $F(1, 38) = 0.97$, $p = .330$, $\eta_p^2 = .03$, that the main effect of time was not significant, $F(2, 76) = 2.18$, $p = .120$, $\eta_p^2 = .05$, and that the interaction between task and time was not significant, $F(2, 76) = 0.58$, $p = .562$, $\eta_p^2 = .02$.

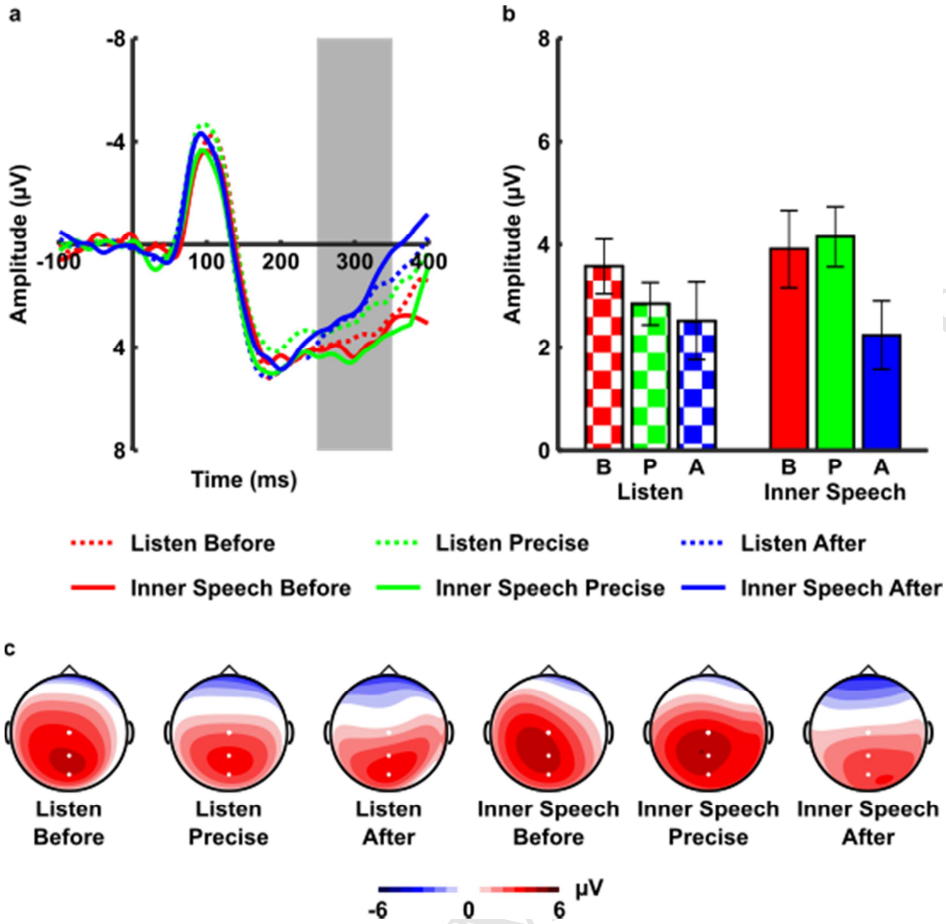


Figure S2. Analysis of the P3. (a) The graph shows the grand-averaged ERPs for each condition averaged over Cz, CPz, and Pz electrodes, showing time (ms) on the *x*-axis, with 0 indicating the onset of the auditory phoneme, and voltage (μV) on the *y*-axis, with negative voltages plotted upwards. The grey bar shows the P3 time-window (250–350 ms), which was selected after visual inspection of the ERPs and voltage maps. (b) The bar graph shows the mean amplitudes for the P3 time-window for the listen and inner speech conditions across the different time delays: before (B), precise (P), and late (L). Error bars show the SEM. (c) The voltage maps show the distribution of voltages over the scalp during the P3 time-window.

Appendix B: Supplementary analyses for Experiment 2

Similar to Experiment 1, we analysed the peak latency of the N1 by identifying the most negative voltage averaged over Fz, FCz, and Cz electrodes in the time-window of 50 to 150 ms for every condition and participant. Repeated-measures ANOVA found a significant main effect of time, $F(2, 108) = 17.41$, $p < .001$, $\eta_p^2 = .24$; however, the main effect of task was not significant, $F(1, 54) = 0.43$, $p = .517$, $\eta_p^2 < .01$, and the interaction between task and time was not significant, $F(2, 108) = 1.58$, $p = .210$, $\eta_p^2 = .03$. Post-hoc *t*-tests found that the peak latency of the N1 was later in the before condition than in the precise, $t(54) = 4.53$, $p < .001$, $d = 0.61$, and after, $t(54) = 4.93$, $p < .001$, $d = 0.67$, conditions. There were no other significant differences.

We then analysed the mean amplitude of the P2 (Crowley & Colrain, 2004) averaged over FCz, Cz, and CPz electrodes in the time-window of 160 to 200 ms. We chose these electrodes and this time-window to be consistent with Experiment 1. Figure S3a shows the ERPs, Figure S3b shows the mean amplitudes for the P2 time-window, and Figure S3c shows the voltage maps for the P2 time-window. Repeated-measures ANOVA found that the main effect of task was not significant, $F(1, 54) = 3.34$, $p = .073$, $\eta_p^2 = .06$, that the main effect of time was not significant, $F(2, 108) = 0.79$, $p = .455$, $\eta_p^2 = .01$, and that the interaction between task and time was not significant, $F(2, 108) = 0.89$, $p = .414$, $\eta_p^2 = .02$.

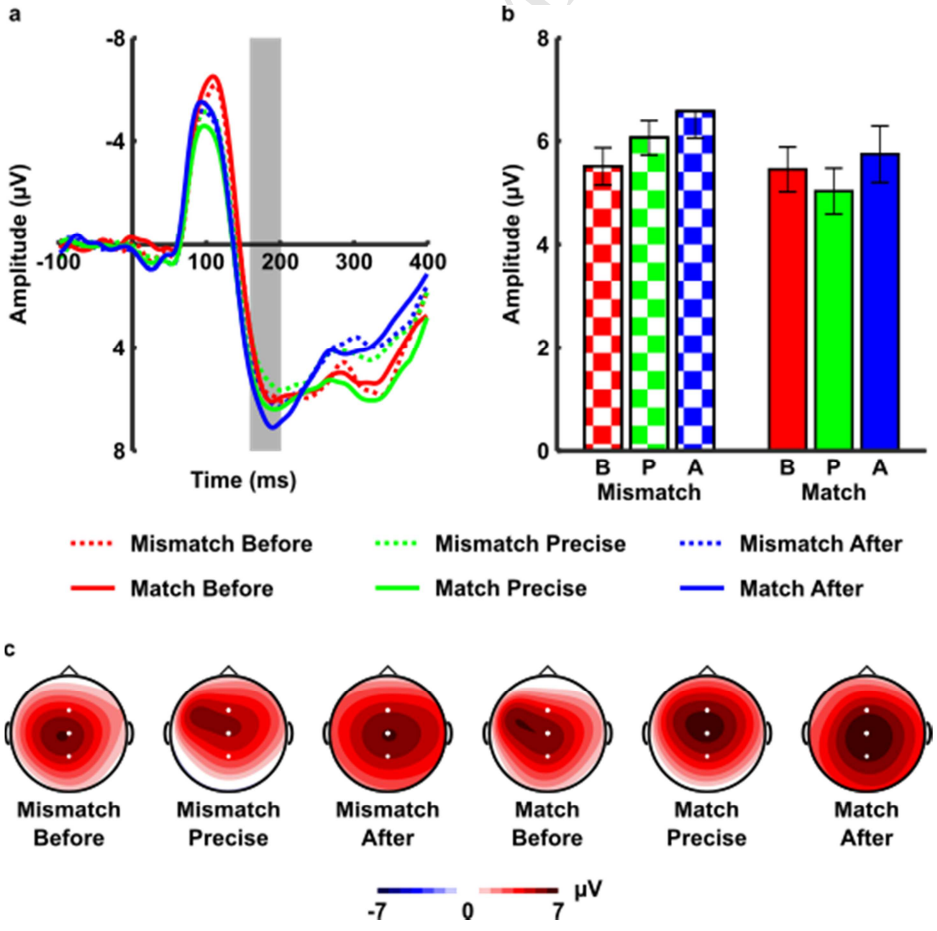


Figure S3. Analysis of the P2. (a) The graph shows the grand-averaged ERPs for each condition averaged over FCz, Cz, and CPz electrodes, showing time (ms) on the *x*-axis, with 0 indicating the onset of the auditory phoneme, and voltage (μV) on the *y*-axis, with negative voltages plotted upwards. The grey bar shows the P2 time-window (160–200 ms), which we used to be consistent with Experiment 1. (b) The bar graph shows the mean amplitudes for the P2 time-window for the listen and inner speech conditions across the different time delays: before (B), precise (P), and late (L). Error bars show the SEM. (c) The voltage maps show the distribution of voltages over the scalp during the P2 time-window.

Finally, we analysed the mean amplitude of the P3 (Polich, 2007) averaged over Cz, CPz, and Pz electrodes in the time-window of 250 to 350 ms. We chose these electrodes and this time-window to be consistent with Experiment 1. Figure S4a shows the ERPs, Figure S4b shows the mean amplitudes for the P3 time-window, and Figure S4c shows the voltage maps for the P3 time-window. Repeated-measures ANOVA found that the main effect of task was not significant, $F(1, 54) = 3.57$, $p = .064$, $\eta_p^2 = .06$, that the main effect of time was not significant, $F(2, 108) = 0.85$, $p = .432$, $\eta_p^2 = .02$, and that the interaction between task and time was not significant, $F(2, 108) = 1.54$, $p = .218$, $\eta_p^2 = .03$.

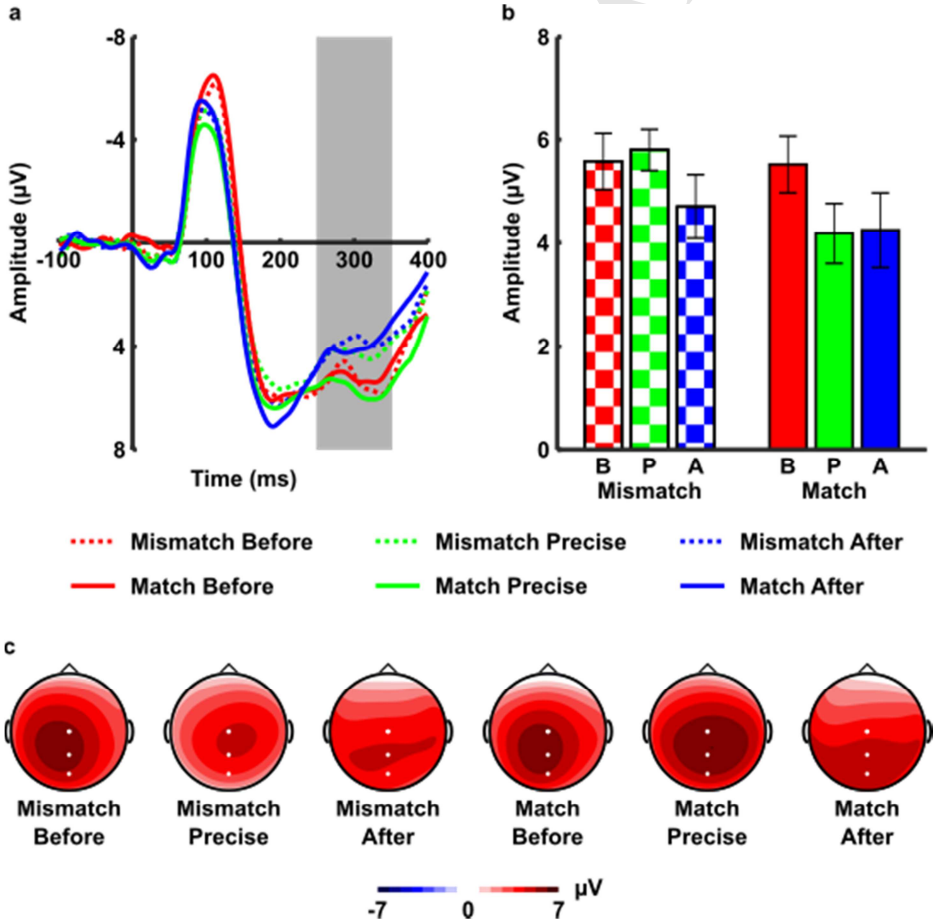


Figure S4. Analysis of the P3. (a) The graph shows the grand-averaged ERPs for each condition averaged over Cz, CPz, and Pz electrodes, showing time (ms) on the *x*-axis, with 0 indicating the onset of the auditory phoneme, and voltage

(μV) on the *y*-axis, with negative voltages plotted upwards. The grey bar shows the P3 time-window (250–350 ms), which we used to be consistent with Experiment 1. (b) The bar graph shows the mean amplitudes for the P3 time-window for the listen and inner speech conditions across the different time delays: before (B), precise (P), and late (L). Error bars show the SEM. (c) The voltage maps show the distribution of voltages over the scalp during the P3 time-window.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*, 931–965.

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition, 65*, 48–58.

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex, 15*, 221–228.

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience, 21*, 791–802.

Bäß, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology, 70*, 137–143.

Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience, 12*, 54: 1-10.

Behroozmand, R., Karvelis, L., Liu, H., & Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clinical Neurophysiology, 120*, 1303–1312.

Behroozmand, R., Liu, H., & Larson, C. R. (2010). Time-dependent neural processing of the auditory feedback during voice pitch error detection. *Journal of Cognitive Neuroscience, 23*, 1205–1217.

Behroozmand, R., Sangtian, S., Korzyukov, O., & Larson, C. R. (2016). A temporal predictive code for voice motor control: Evidence from ERP and behavioral responses to pitch-shifted auditory feedback. *Brain Research, 1636*, 1–12.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience, 1*, 635–640.

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Chen, Z., Chen, X., Liu, P., Huang, D., & Liu, H. (2012). Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neuroscience, 13*, 55: 1–10.

Crapse, T. B., & Sommer, M. A. (2008). Corollary discharge circuits in the primate brain. *Current Opinion in Neurobiology, 18*, 552–557.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*, 732–744.

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature, 453*, 1102–1106.

Elijah, R. B., Le Pelley, M. E., & Whitford, T. J. (2016). Modifying temporal expectations: Changing cortical responsivity to delayed self-initiated sensations with training. *Biological Psychology*, *120*, 88–95.

Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, *4*, 636–640.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*, 48–58.

Ford, J. M., & Mathalon, D. H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, *38*, 37–46.

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*, *17*, 631–648.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*, 180–190.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*, 1125–1138.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, *16*, 391–399.

Kilteni, K., Andersson, B. J., Houborg, C., & Ehrsson, H. H. (2018). Motor imagery involves predicting the sensory consequences of the imagined movement. *Nature Communications*, *9*, 1617: 1-9.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Supplement), 14.

Knolle, F., Schröger, E., & Kotz, S. A. (2013). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, *92*, 410–416.

Lebedev, M. A., & Nicolelis, M. A. (2006). Brain-machine interfaces: Past, present and future. *Trends in Neurosciences*, *29*, 536–546.

Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*, *122*, 951–957.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*, 146–157.

Mellor, C. S. (1970). First rank symptoms of schizophrenia. *The British Journal of Psychiatry*, *117*, 15–23.

Miall, R. C., & Wolpert, D. (1996). Forward models for physiological motor control. *Neuronal Networks*, *9*, 1265–1279.

Mifsud, N. G., Oestreich, L. K. L., Jack, B. N., Ford, J. M., Roach, B. J., Mathalon, D. H., & Whitford, T. J. (2016). Self-initiated actions result in suppressed auditory but amplified visual evoked components in healthy participants. *Psychophysiology*, *53*, 723–732.

Miller, G. A., Gratton, G., & Yee, C. M. (1988). Generalized implementation of an eye movement correction procedure. *Psychophysiology*, *25*, 241–243.

Morin, A., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, *32*, 376–382.

Morin, A., Uttl, B., & Hamper, B. (2011). Self-reported frequency, content, and functions of inner speech. *Social and Behavioural Sciences*, *30*, 1714-1718.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Oestreich, L. K. L., Mifsud, N. G., Ford, J. M., Roach, B. J., Mathalon, D. H., & Whitford, T. J. (2016). Cortical suppression to delayed self-initiated auditory stimuli in schizotypy: Neurophysiological evidence for a continuum of psychosis. *Clinical EEG and Neuroscience*, *47*, 3–10.

Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *NeuroImage*, *14*, 182–193.

Pelli, D. G. (1997). The Videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, *261*, 220–239.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148.

SanMiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP. *Psychophysiology*, *50*, 334–343.

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological Science*, *24*, 1824–1830.

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, *31*, 241–253.

Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, *93*, 20–31.

Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PLoS ONE*, *8*, e82925: 1–8.

Sokolov, A. N., Onischenko, G. T., & Lindsley, D. B. (1972). *Inner speech and thought*. New York: Plenum Press.

Sperry, R. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology, 43*, 482–489.

Straka, H., Simmers, J., & Chagnaud, B. P. (2018). A new perspective on predictive motor signaling. *Current Biology, 28*, R193–R194.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*, 166: 1-23.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*, 314: 1-11.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*(3), 1020–1036.

Tian, X., & Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience*, *27*, 352–364.

Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour*, *2*, 225–234.

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1–12.

Timm, J., SanMiguel, I., Saupe, K., & Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience*, *14*, 2: 1–11.

Virtanen, J., Ahveninen, J., Ilmoniemi, R. J., Näätänen, R., & Pekkonen, E. (1998). Replicability of MEG and EEG measures of the auditory N1/N1m-response. *Electroencephalography and Clinical Neurophysiology*, *108*, 291–298.

Von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip [The reafference potential]. *Naturwissenschaften*, *37*, 464–476.

Whitford, T. J., Ford, J. M., Mathalon, D. H., Kubicki, M., & Shenton, M. E. (2012). Schizophrenia, myelination, and delayed corollary discharges: A hypothesis. *Schizophrenia Bulletin*, *38*, 486–494.

Whitford, T. J., Jack, B. N., Pearson, D., Griffiths, O., Luque, D., Harris, A. W. F., … Le Pelley, M. E. (2017). Neurophysiological evidence of efference copies to inner speech. *eLife*, *6*, e28197: 1–23.

Whitford, T. J., Mathalon, D. H., Shenton, M. E., Roach, B. J., Bammer, R., Adcock, R. A., … Ford, J. M. (2011). Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychological Medicine*, *41*, 959–969.

Woods, D. L. (1995). The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology*, *44*, 102–109.

World Medical Association. (2004). Declaration of Helsinki: Ethical principles for medical research involving human subjects.

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., et al. (2015). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cerebral Cortex*, *25*, 1576–1586.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, *8*, 29–46.

Zivin, G. (1979). *The development of self-regulation through private speech*. New York: Wiley.