Inner speech captures the perception of external speech

Mark Scott[a)]
Linguistics Department, University of British Columbia, Vancouver, Canada
shark_scott@hotmail.com

H. Henny Yeung[b)]
Psychology Department, University of British Columbia, Vancouver, Canada
henny.yeung@parisdescartes.fr

Bryan Gick[c)]
Linguistics Department, University of British Columbia, Vancouver, Canada
gick@interchange.ubc.ca

Janet F. Werker
Psychology Department, University of British Columbia, Vancouver, Canada
jwerker@psych.ubc.ca

Abstract: Talking silently to ourselves occupies much of our mental lives, yet the mechanisms underlying this experience remain unclear. The following experiments provide behavioral evidence that the auditory content of inner speech is provided by corollary discharge. Corollary discharge is the motor system's prediction of the sensory consequences of its actions. This prediction can bias perception of other sensations, pushing percepts to match with prediction. The two experiments below show this bias induced by inner speech, demonstrating that inner speech causes external sounds to be heard as similar to the imagined speech, and that this bias operates on subphonemic content.

PACS numbers: 43.66.Ba, 43.70.Mn, 43.71.An [QJF]
Date Received: November 28, 2012 **Date Accepted:** February 26, 2013

1. Introduction

Throughout the day most of us engage in a nearly ceaseless internal banter. This stream of inner speech is a core aspect of our mental lives, and is linked to a wide array of psychological functions. Despite this centrality, inner speech has received little scientific attention.[1] The experiments below attempt to redress this by providing behavioral evidence that the auditory content of inner speech (the "sound" of the voice in your head) is provided by corollary discharge.

Corollary discharge is a neural signal generated by the motor system that serves to prevent confusion between self-caused and externally-caused sensations. When an animal performs an action, its motor system uses a *forward model* (an internal model of the body) to predict the sensory consequences that will result. This prediction is *corollary discharge.* Corollary discharge is relayed to sensory areas where it serves to segregate out incoming sensations that match the prediction since these are

[a)]Author to whom correspondence should be addressed. Current address: Linguistics Department, Kobe Shoin University, Kobe, Japan.
[b)]Current address: Laboratoire Psychologie de la Perception (CNRS UMR 8158) and Université Paris Descartes, Paris, France.
[c)]Also at: Haskins Laboratories, New Haven, CT.

likely to be caused by the animal's action.[2] Corollary discharge is what allows you to speak without confusing your voice with other voices/sounds in the environment.[3]

Auditory corollary discharge (for speech) is therefore an internal prediction of the sound of one's own voice. The similarity to inner speech is striking and motivates the theory, tested below, that the sound of inner speech is constituted by corollary discharge. The experiments below contribute behavioral evidence to a growing literature on this topic, including recent papers by Tian and Poeppel,[4] Scott,[5] Pickering and Garrod,[6] and Grush.[7] While brain-imaging studies have already provided some support for this theory, showing that one indication of corollary discharge, known as *sensory attenuation*, is induced by inner speech,[8,9] behavioral evidence has been lacking. The current pair of experiments examine a behavioral hallmark of corollary discharge: *Perceptual capture*.

Perceptual capture is a shift in perception caused by the fact that corollary discharge is an anticipation and as such can pull ambiguous stimuli into alignment with the anticipated percept. Repp and Knoblich demonstrated such a capture effect by having pianists perform a hand motion that was consistent with playing either a rising or falling sequence of tones.[10] These hand motions were performed in time to a tonal sequence that was specially constructed to be perceptually ambiguous between ascending and descending. When performing an ascending hand motion, pianists tended to hear the ambiguous sound as ascending. Schütz-Bosbach and Prinz review a large number of perceptual capture effects across several sensory modalities,[11] and Hickok *et al.* provide an overview of how corollary discharge influences perception through its role as an anticipation.[12]

The following experiments look for evidence of perceptual capture from inner speech, which would support the theory that the auditory content of inner speech is provided by corollary discharge. Sams *et al.* have already demonstrated that silent articulation of syllables (which they assume engages auditory corollary-discharge) induces perceptual capture.[13] While Sams *et al.* do not relate this to inner speech, silent articulation (mouthing) *is* a form of inner speech. In fact, inner speech exists on a continuum of articulator engagement: From "mouthing," where the articulators are moving fully (but silently), to "pure imagery" where the articulators remain still.[14]

The strength of perceptual capture is not predicted to be equivalent in these two forms of inner speech, however. Because the motor system generates corollary discharge, we expect that greater motor engagement should trigger corollary discharge more strongly, predicting that perceptual capture should be stronger in mouthed than in pure speech imagery. Oppenheim and Dell make a similar claim, arguing that the subphonemic content of inner speech is positively correlated with articulator engagement.[14] Both experiments below test this claim and find stronger perceptual capture in mouthed than pure inner speech.

In Experiment 1 we extend the work of Sams *et al.*, showing that pure inner speech, not just mouthed inner speech, can cause an ambiguous speech sound to be heard as similar to the imagined phoneme. We argue that this is due to corollary discharge. Since corollary discharge is not a phonemic representation, this means that phonemic identity between the imagined speech sound and resulting percept (as in Experiment 1) should not be necessary, and perceptual capture should also occur when the sounds are merely similar. Experiment 2 demonstrates this, showing that an ambiguous sound is heard as *more similar* to an imagined sound, even when they belong to different phonological categories.

Oppenheim and Dell also look for effects in inner speech of content at a finer level of detail than the phoneme, referring to this as "subphonemic" content.[14] It is an open question whether this subphonemic content consists of a sensory or phonological (featural) representation. We follow Oppenheim and Dell in referring to subphonemic content in inner speech, keeping in mind that the corollary discharge literature would refer to this content as "sensory." The degree to which sensory, phonological, and motor codes are related is an open question which we do not address here.

2. Experiment 1

Participants were asked to mouth or imagine (n.b., "imagine" will be used hereafter to mean producing "pure," non-articulated, inner speech) /ɑˈvɑ/ or /ɑˈbɑ/ in synchrony with an external sound that was itself ambiguous between /ɑˈvɑ/ and /ɑˈbɑ/, and then to categorize the ambiguous sound. The prediction is that when mouthing or imagining a sound, the categorization of external ambiguous sounds will be biased toward the content of inner speech. These experiments extend the research of Sams *et al.*, replicating the effects of mouthing speech sounds and extending them to pure imagery as well.

2.1 Methods

Experiment 1 consisted of 5 blocks: **Mouth /ɑˈbɑ/, Imagine /ɑˈbɑ/, Mouth /ɑˈvɑ/, Imagine /ɑˈvɑ/** and **Baseline** (in which participants categorized the targets without performing any form of inner speech). Participants cycled through these blocks, in random order, performing each block 14 times, with 6 trials per block.

In order to ensure synchrony between inner and external sounds, a rhythm was established by playing a "murmured" sound prior to target presentation. The murmured sounds were low-pass filtered versions of the targets, filtered at 275 Hz so only pitch, timing, and amplitude information were available. One token of the murmur was played prior to each target so participants could align their mouthing/imagining to its timing. To augment the sense of rhythm, a video of a red dot (like in karaoke) flashed in time to both murmur and target sounds.

Speech-like murmur sounds were used to establish the rhythm because synchronizing with speech sounds is a familiar activity (e.g., lip synching to songs). The filtering used, 275 Hz, should be sufficient to eliminate all phonetically informative information. However, even if some small amount of information survived the filtering, the same murmur and target sounds were used in all conditions, so differences between conditions cannot be attributed to priming from these murmur sounds.

On each trial, participants mouthed/imagined twice (except in the **Baseline** condition, in which they merely listened), once in time to the murmur and then once in time to an ambiguous target sound. After mouthing/imagining, participants categorized the target.

There were 20 female participants [average age = 21.7 ys; standard deviation (SD) = 3.7]. Female participants were used so that the stimuli and participants were of the same sex (to ensure that there would be no gender conflict between corollary discharge and heard stimuli).

A 168-step female-voice continuum from /ɑˈbɑ/ to /ɑˈvɑ/ was created using STRAIGHT.[15] A pre-test with probit analysis was performed before the main experiment to determine each participant's 44%, 50%, 56% points in perceptual space between /ɑˈbɑ/ and /ɑˈvɑ/. These points were used as the ambiguous targets. Three points were used simply to provide variation in the target stimuli, so that participants did not feel they were always categorizing the same sound.

2.2 Results

Responses to the three targets were pooled for analysis. A repeated-measures analysis of variance (ANOVA) was performed with the five types of experimental block as five levels of a single factor. This was highly significant [$F(4,76) = 39.489$, $p < 0.001$]. Six planned comparisons were made, using pairwise t-tests with Holm-Bonferroni correction. These confirmed that each of the experimental conditions (**Mouth /ɑˈbɑ/, Imagine /ɑˈbɑ/, Mouth /ɑˈvɑ/** and **Imagine /ɑˈvɑ/**) was significantly different from **Baseline** ($p < 0.05$). As well, **Mouth /ɑˈbɑ/** was significantly different from **Imagine /ɑˈbɑ/** ($p < 0.001$) and likewise **Mouth /ɑˈvɑ/** was significantly different from **Imagine /ɑˈvɑ/** ($p < 0.001$), showing that the impact of mouthing was stronger than that of pure imagery, as predicted. These results are shown in Fig. 1.

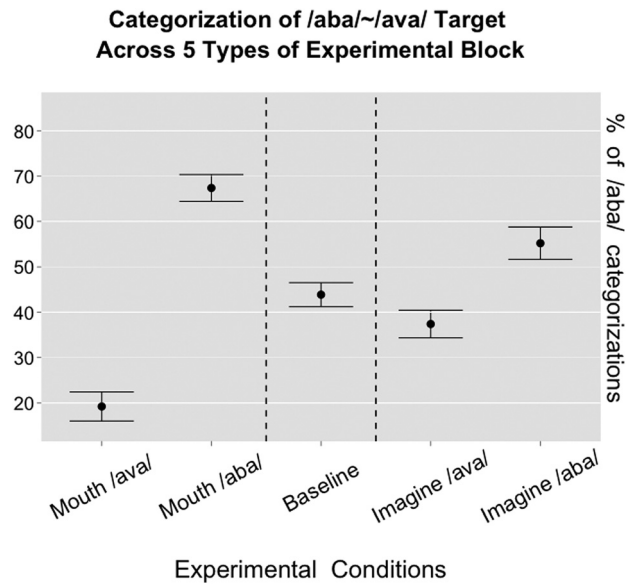

Fig. 1. Experiment 1 results: The data are scored as percent of targets categorized as /ɑˈbɑ/. Standard error bars are shown.

2.3 Discussion

For both mouthed and pure inner speech, participants were more likely to hear an ambiguous sound as matching the content of their imagery. The two-way directionality of the effect (mouthing/imagining /ɑˈbɑ/ pulling perception in one direction but mouthing/imagining /ɑˈvɑ/ pulling in the opposite direction) demonstrates that it is the content of the inner speech that is responsible for the effect, not some extraneous factor.

This experiment also shows a distinction between **Mouthing** and **Imagining:** For both /ɑˈbɑ/ and /ɑˈvɑ/, the **Mouth** conditions were significantly different from the corresponding **Imagine** conditions. This is predicted under the assumption that greater articulator engagement triggers more corollary discharge engagement.

3. Experiment 2

Experiment 1 demonstrated perceptual capture from inner speech—an ambiguous external sound was perceived to be the same phoneme as an imagined sound. If this perceptual capture is caused by corollary discharge, as we claim, then it is not due to the imagined sound being the same phoneme as the percept, i.e., it is not simply a case of phoneme priming.

To test this, Experiment 2 demonstrates that inner speech can make the perception of an external sound match the subphonemic aspects of an imagined sound, *without* there being phonemic identity between imagined sound and percept. By extension, this experiment also tests the claim that subphonemic content exists in inner speech.

In addition, as with Experiment 1, mouthed and pure inner speech are compared. In their experiments Oppenheim and Dell demonstrate that subphonemic similarity induces speech errors in mouthed inner speech, but their experimental setup did not have sufficient power to demonstrate an effect of subphonemic similarity in pure inner speech. Experiment 2, using a different experimental procedure, succeeds in showing that subphonemic content, while weaker (as predicted), is still present in pure inner speech.

3.1 Methods

This experiment was identical to Experiment 1; only the sounds being mouthed/imagined were different. The mouthed/imagined sounds were /ɑˈfɑ/ and /ɑˈpɑ/. /ɑˈfɑ/ is

similar to /ɑ'vɑ/ at a subphonemic level (both are labiodental fricatives) and /ɑ'pɑ/ is similar to /ɑ'bɑ/ (both are bilabial stops). If imagery of /ɑ'fɑ/ can cause an /ɑ'bɑ/ ~ /ɑ'vɑ/ ambiguous target to be perceived as /ɑ'vɑ/ (and the converse for imagery of /ɑ'pɑ/) that would indicate the presence of perceptual capture at the subphonemic level. This was tested in both mouthed and pure inner speech.

There were 20 new female participants (average age = 21.5 ys; SD = 3.5).

The same stimuli were used as in Experiment 1.

3.2 Results

Responses to the three targets were pooled for analysis. A repeated-measures ANOVA was performed with the five types of experimental block as five levels of a single factor. This was highly significant [$F(4,76) = 52.215$, $p < 0.001$]. Six planned comparisons were made (using pairwise *t*-tests with Holm-Bonferroni correction). These confirmed that each of the experimental conditions (**Mouth /ɑ'pɑ/, Imagine /ɑ'pɑ/, Mouth /ɑ'fɑ/** and **Imagine /ɑ'fɑ/**) was significantly different from **Baseline** ($p < 0.05$). As well, **Mouth /ɑ'pɑ/** was significantly different from **Imagine /ɑ'pɑ/** ($p = 0.0173$) and likewise **Mouth /ɑ'fɑ/** was significantly different from **Imagine /ɑ'fɑ/** ($p < 0.001$), showing that the impact of mouthing was stronger than that of pure imagery. These results are shown in Fig. 2.

3.3 Discussion

As with Experiment 1, imagery influenced the perception of ambiguous /ɑ'bɑ/ ~ /ɑ'vɑ/ targets. Mouthing or imagining /ɑ'fɑ/ biased perception towards /ɑ'vɑ/, and conversely, mouthing or imagining /ɑ'pɑ/ biased perception toward /ɑ'bɑ/. This shows that the perceptual capture found in Experiment 1 can be induced at the subphonemic level. Furthermore, unlike the experiments of Oppenheim and Dell,[14] this experiment succeeded in showing subphonemic influences from pure inner speech in addition to mouthed inner speech.

As with Experiment 1, this experiment showed that mouthed inner speech has a greater impact on perception than does pure inner speech—which is predicted under a corollary discharge account and supports the claim of Oppenheim and Dell that

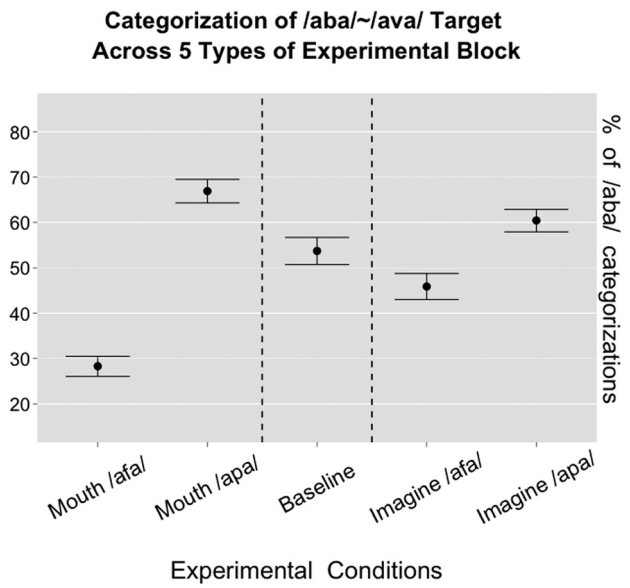

Fig. 2. Experiment 2 results: The data are scored as percent of targets categorized as /ɑ'bɑ/. Standard error bars are shown.

subphonemic content in inner speech exists on a continuum tied to articulator engagement.[14]

4. General discussion and conclusions

Corollary discharge constitutes an anticipation of hearing a particular sound, and as such tends to capture incoming sensations. If speech imagery involves corollary discharge, this capture effect should be induced by inner speech. These two experiments successfully tested this prediction: Inner speech (both mouthing and pure) influenced perception, biasing participants to hear ambiguous sounds as similar to the content of their imagery.

In Experiment 1 people heard more /ɑˈbɑ/ when mouthing/imagining /ɑˈbɑ/ and more /ɑˈvɑ/ when mouthing/imagining /ɑˈvɑ/. This replicates the perceptual capture effect reported in Sams *et al.* for mouthing,[13] and extends the effect to pure inner speech.

Experiment 2 shows that this perceptual capture is not dependent on phonemic identity between imagined sound and percept, as the influence still occurs when the sounds are merely similar. This demonstrates that the perceptual capture found in Experiment 1 is not simply a matter of phoneme priming. It also shows that inner speech does indeed contain information below the level of the phoneme. Whether this information is a phonological representation or a more sensory one is a question for further research.

The perceptual capture demonstrated in these experiments is related to recent theoretical and experimental work which suggests that forward models and corollary discharge can play a role in normal speech perception, by providing top-down information that constrains potential perceptual interpretations, thus lowering the chance of error.[4] This function of forward models in speech perception may be viewed as similar to the Motor Theory of Speech Perception.[16] The primary difference is that motor-system involvement is viewed as obligatory by the Motor Theory, while it is viewed as a supplementary perceptual strategy in forward-model frameworks. The experiments reported here do not distinguish between these possibilities.

In both experiments, the impact of mouthed inner speech was stronger than that of pure inner speech. This is expected if the auditory content of inner speech is provided by corollary discharge, which should be stronger with greater motor involvement.

Together, these experiments provide support for the claim that the inner speech has rich auditory content (including, at a minimum, subphonemic information) that is provided by corollary discharge.

Acknowledgments

We would like to acknowledge the financial support of NSERC *Discovery Grants* to B.G. and J.F.W.

References and links

[1]R. Jackendoff, *Consciousness and the Computational Mind* (MIT Press, Cambridge, MA, 1987).
[2]S. J. Eliades and X. Wang, "Neural substrates of vocalization feedback monitoring in primate auditory cortex," Nature **453**(7198), 1102–1106 (2008).
[3]S. O. Aliu, J. F. Houde, and S. S. Nagarajan, "Motor-induced suppression of the auditory cortex," J. Cogn Neurosci. **21**(4), 791–802 (2009).
[4]X. Tian and D. Poeppel, "Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation," Front. Human Neurosci. **6**(314), 1–11 (2012).
[5]M. Scott, "Corollary discharge provides the sensory content of inner speech," Psychol. Sci. (in press).
[6]M. J. Pickering and S. Garrod, "An integrated theory of language production and comprehension," Behav. Brain Sci. (in press).
[7]R. Grush, "The emulation theory of representation: motor control, imagery, and perception," Behav. Brain Res. **27**(3), 377–396; discussion 396–442 (2004).

[8]J. Kauramäki, I. P. Jääskeläinen, R. Hari, R. Möttönen, J. P. Rauschecker, and M. Sams, "Lip reading and covert speech production similarly modulate human auditory-cortex responses to pure tones," J. Neurosci. **30**(4), 1314–1321 (2010).
[9]X. Tian and D. Poeppel, "Mental imagery of speech and movement implicates the dynamics of internal forward models," Front. Psychol. **1**, 1–23 (2010).
[10]B. H. Repp and G. Knoblich, "Performed or observed keyboard actions affect pianists' judgments of relative pitch," Q. J. Exp. Psychol. A **62**(11), 2156–2170 (2009).
[11]S. Schütz-Bosbach and W. Prinz, "Perceptual resonance: Action-induced modulation of perception," Trends Cogn. Sci. **11**(8), 349–355 (2007).
[12]G. Hickok, J. Houde, and F. Rong, "Sensorimotor integration in speech processing: computational basis and neural organization," Neuron **69**(3), 407–422 (2011).
[13]M. Sams, R. Möttönen, and T. Sihvonen, "Seeing and hearing others and oneself talk," Brain Res. Cognit. Brain Res. **23**(2–3), 429–435 (2005).
[14]G. M. Oppenheim and G. S. Dell, "Motor movement matters: The flexible abstractness of inner speech," Mem. Cognit. **38**(8), 1147–1160 (2010).
[15]H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous frequency-based f0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207 (1999).
[16]A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," Cognition **21**(1), 1–36 (1985).