Breathing position influences speech perception

Mark Scott (mark.a.j.scott@gmail.com)
Department of Linguistics, United Arab Emirates University
Al Ain, Abu Dhabi

Henny Yeung (henny.yeung@parisdescartes.fr)
Laboratoire Psychologie de la Perception (UMR 8242)
CNRS & Université Paris Descartes, Paris, France

Abstract

Participants were asked to breath through their mouth or their nose, forcing them to adopt a particular position of their velum (up or down). While breathing in each of these positions, they categorized sounds from an /ada/ to /ana/ continuum. The position of the speech articulators, even though adopted for the purposes of breathing, altered participants' perception of external speech sounds, so that when the velum was down (to breath through the nose), they tended to hear the consonant as the nasal /n/ — a sound necessarily produced with a lowered velum, rather than as /d/ in which the velum must be raised.

Keywords: speech perception; articulation; corollary discharge; efference copy; forward model; motor control; breathing; motor theory

Introduction

If the organs of speech production are moved during speech perception, that movement can influence how perception unfolds. For example, if hearing a sound ambiguous between /aba/ and /ava/ while mouthing (or even imagining) /ava/, a person will tend to hear the sound as /ava/ (and contrariwise for mouthing /aba/) (Scott, Yeung, Gick, & Werker, 2013). This phenomenon is purportedly due to the perceptual anticipation (in the form of *corollary discharge*) caused by mouthing/imagining. This experiment explores whether a similar perceptual capture can occur even when the organs of speech production are not being engaged in a speech task — specifically, whether the position of the velum (up for breathing through the mouth, down for breathing through the nose) can influence the perception of nasal vs. non-nasal stop consonants.

Motor involvement in speech perception

The question of whether (or to what degree) the motor system influences speech perception is old and strongly contested. The best known version of this idea is the Motor Theory of Speech Perception (Liberman & Mattingly, 1985) which claims that speech perception is achieved via a specialized module which extracts the intended speech gestures from the acoustic signal.

A somewhat similar view is proposed by Fowler (1986) who offers a Gibsonian approach to speech perception in which it is speech gestures that are recovered in perception but, in contrast to the Motor Theory, this recovery is through general auditory mechanisms rather than a biologically specialized speech-perception mechanism. The view that the fundamental units of speech perception (and production) are gestures is also shared by the Gestural Phonology approach (Goldstein & Fowler, 2003).

More recent alternatives to the Motor Theory (and its variants) maintain a role for the motor system, but do not claim that it is an obligatory component of speech perception. For example Pickering and Garrod (2007), propose that when perception is faced with a difficult task, top-down information in the form of motor-predictions can be used to 'fill in the gaps'. According to this theory, speech perception would primarily be an auditory process but when the auditory signal is particularly unclear the motor system makes predictions about what is to come and, in so doing, constrains the possibilities that the auditory system must entertain, thus easing the computational load. This would be a hybrid motor/auditory view of speech perception. Skipper, Nusbaum, and Small (2006) and Skipper, van Wassenhove, Nusbaum, and Small (2007) have proposed a similar theory, again arguing that speech perception is not necessarily a matter of coding the incoming sound into a gestural code, but that engagement of predictions from the motor system can be used to aid in speech-perception. Skipper et al. (2007) argue that such predictions are what underlie the influence of vision on speech perception, such as in the McGurk effect (McGurk & MacDonald, 1976).

These motor-helping-hearing theories are quite similar to the Perception-for-Action-Control Theory (Schwartz, Basirat, Ménard, & Sato, 2010). This theory argues that speech perception is not motor-based, but that speech gestures do define equivalence classes for speech sounds. The idea is that the motor system helps establish which sounds count as members of the same category, membership being determined by sharing a common method of production. However, once the sound classes are set, the motor system is normally *not* used online in the act of perceiving the members of these classes, such online perception being achieved by the auditory system. Schwartz et al. (2010) argue for one exception to the independence of sensory and motor processes — when auditory perception is made difficult because of missing information, the motor system can be used to 'fill in' that missing information.

Corollary Discharge

These recent alternatives to the Motor Theory propose a specific mechanism by which the motor system influences speech perception: corollary discharge.[1]

Corollary discharge is an internal sensory signal generated by one's own motor system whenever one acts (Aliu, Houde, & Nagarajan, 2009). One primary function of corollary discharge is to provide *pseudo* feedback in situations where regular sensory feedback is too slow to guide one's actions. There is an unavoidable time delay in real sensory feedback — our senses do not operate instantaneously, it takes time for a change in the environment (or in our body) to be transduced by our end-organs, then transmitted to and processed by the central nervous system and for a motor correction on the basis of this information to be issued. This delay can be quite considerable – for auditory speech perception it has been estimated at around 130 ms (Jones & Munhall, 2002). This means that feedback is not available (or minimally available) for speech movements that are faster than 130ms (which is in fact many speech sounds). Corollary discharge can be generated by the motor system *before* sensory feedback is available and thus can serve the feedback role and avoid the time-lag problem (Wolpert & Flanagan, 2001).

Another role of corollary discharge is to tag concurrent matching external sensations as "self-caused" and so unworthy of intense perceptual processing (Eliades & Wang, 2008). This function means that corollary discharge functions to anticipate perceptions and so can pull ambiguous stimuli into alignment with the anticipated percept. Scott et al. (2013) hypothesized that corollary discharge, generated during mouthing of speech, channels the perception of external speech into matching the corollary discharge prediction, thus performing a *perceptual capture* function.

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. Hickok, Houde, and Rong (2011) provide an overview of how corollary discharge influences perception through its role as an anticipation.

Repp and Knoblich (2009) demonstrated perceptual capture from motor-induced anticipations by having pianists perform hand motions for a rising or falling sequence of notes. These hand motions were performed in synchrony with a sequence of notes that could be heard (thanks to a perceptual illusion) as rising or falling. When performing the hand motion consistent with a rising sequence, pianists tended to hear the ambiguous sound as ascending. Schütz-Bosbach and Prinz (2007) review several such perceptual capture effects across sensory modalities.

In terms of hearing, several studies have found evidence of anticipations altering the perception of sounds. A compelling example of this is the "White Christmas" effect (Merckelbach & Ven, 2001) in which people are induced to hear the song "White Christmas" when presented with white noise, simply by telling them that the song *might* be buried under the noise (but is not). Such perceptual shifts in speech perception arising from the influence of the motor system have been shown in other studies such as Sams, Möttönen, and Sihvonen (2005), Ito, Tiede, and Ostry (2009) and Scott et al. (2013).

A consequence of theories which propose a perceptual 'fill in the gap' role for corollary discharge is that the position of the perceiver's own articulators should matter for what information gets filled in. For a prediction of the sensory consequences of an action to be accurate it must take into account the starting point of the action, as the sensory consequences of an action can be vastly different depending on where the effector is starting — think of the tactile sensory difference between slamming your jaw shut when your tongue is in its normal resting position vs. extended out between your teeth (ouch!). Thus corollary discharge is necessarily generated using the current position of the articulators as the basis for prediction (Houde & Nagarajan, 2011; Hickok, 2012).

Prediction

Given that the predictions of corollary discharge take into account the position of the effectors, the perceptual channelling discussed above should be sensitive to the positions of one's own speech articulators (even if one is not speaking). Thus when a person's velum is down in order to allow breathing through the nose, then the perceptual channelling (or 'filling in') done by the motor system should be biased towards a prediction of nasality and the person should thus be more likely to hear an ambiguous external sound as nasal. In the context of this experiment, this means that people should hear an /ada/~/aga/ ambiguous sound more often as /ana/ when they are breathing through their nose in comparison to when they are breathing through their mouth.

The sounds /d/ and /n/ were chosen as these sounds have the same place of articulation and are both voiced and are both stops, differing almost exclusively in whether the velum is up (for /d/ and so no airflow through the nose) or down (for /n/ with airflow through the nose). Thus the primary difference between these sounds is mirrored in the position of the velum for breathing — up for breathing through the mouth, down for breathing through the nose.

The prediction that people should hear more /n/ when their velum is down for breathing through their nose is similar to the perceptual capture effect demonstrated in Sams et al. (2005) or Scott et al. (2013) but, unlike those experiments, in the current experiment the articulators are not being used in a speech task by the perceiver.

This experiment is also similar to that of Ito et al. (2009), in which they showed that dynamic deformation of a perceiver's face (by a robot device) can alter perception in line with the movement, but only if the movement is timed appropriately with the percept. In contrast, the current experiment asks

[1] Corollary discharge is the sensory prediction generated by a 'forward model' – which is a system that takes motor commands as input and predicts the sensory consequences. Efference copy refers to the motor command received by the forward model, but sometimes the terms *corollary discharge* and *efference copy* are used interchangeably.

whether a static articulator position can also induce a shift in perception.

Methods

Participants were asked to breath through their mouth or nose while they categorized sounds as /ada/ or /ana/. The prediction is that when participants are breathing through their nose, the necessarily lowered velum will influence their perception so that they hear the sounds as more similar to the nasal /ana/. In a control experiment, participants categorized /ada/ vs. /aga/ while breathing through their mouth or nose. No difference was predicted for this control experiment.

Stimuli

A female native speaker of standard European French was recorded saying /ada/ and /ana/. A 10 000 step continuum between these sounds was created using STRAIGHT (Kawahara et al., 2008; Kawahara, Irino, & Morise, 2011). While this may seem like a large number of continuum steps, it should be kept in mind that participants only heard a small subset of these sounds and the large number of steps is simple to generate and allows for very fine-grained precision in estimating phoneme boundaries.

Procedures

There were two conditions:

1. **Breathing through the mouth (velum up)**
2. **Breathing through the nose (velum down)**

In each condition, stimuli were presented using the staircase method (Cornsweet, 1962). The staircase method presents points along a continuum and shifts the subsequent target for presentation based on previous responses. Thus it is able to 'search out' the perceptual boundary between sounds quite quickly. Two interleaved staircases with random switching (to prevent participants being able to predict the upcoming sound) were used for each condition, and participants alternated back and forth between conditions so that each participant performed the interleaved staircase procedure for each condition twice (a total of four staircases per condition). Thus the experiment determined each participant's perceptual boundary between /ada/ and /ana/ while participants breathed through their mouth or nose.

Each staircase consisted of thirteen reversals with decreasing stepsize after each reversal. The step sizes were: 1250, 1000, 800, 650, 500, 350, 250, 150, 80, 50, 30, 20, 10 (from a 10 000 step continuum). The two interleaved staircases started at points 2400 and 7600 of the continuum.

An abstract example of the layout of a staircase procedure is shown in Figure 1. A sound from one end of the continuum (for example the /ada/ end of the continuum) is played to the participant. If the participant categorizes it as belonging to the category consistent with that end of the continuum (/ada/ in our example), then the computer selects the next sound to be closer to the other (/ana/) end of the continuum. If the participant hears this also as /ada/, the computer chooses the next sound to be even closer to /ana/ and so on until the participant reports hearing /ana/. At this point, the computer reverses direction of sound selection (hence this is called a 'reversal') and selects the next sound to be closer to the /ada/ end of the continuum, however the stepsize along the continuum is made smaller (the computer makes smaller jumps along the continuum between sound selections, so that there is more precision). When the participant starts to hear the sound as /ada/ again, the computer reverses again (and again makes the stepsizes along the continuum smaller) moving back toward /ana/. This back and forth continues as the computer homes in on the participant's boundary between /ada/ and /ana/, changing direction of movement along the continuum and getting more precise with each reversal. This is a robust and relatively quick method of estimating a person's perceptual boundary between two sounds.

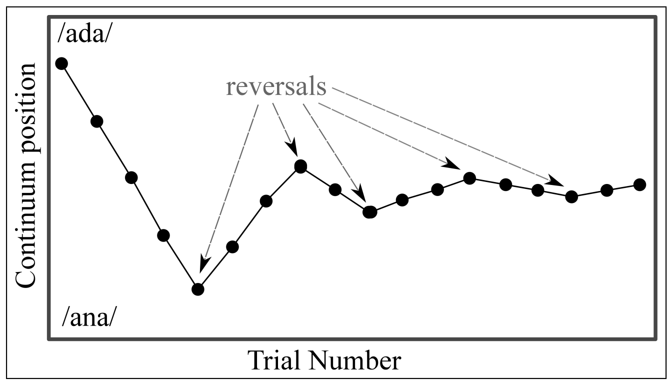

Figure 1: Abstract Example of a Staircase Procedure

The structure of each trial was very simple, participants were presented with an audio stimulus that was somewhat ambiguous between /ada/ or /ana/ and they pressed (with their right hand) a keyboard button (right or left arrow key) to indicate their perception of the sound as /ada/ or /ana/.

Participants were given instructions on the task and familiarized with the software before conducting the experiment. The experiment itself took about 20 minutes for each participant to complete.

The order of breathing conditions (half the participants starting with breathing through the mouth half starting with breathing through the nose) was counterbalanced across participants as was the correspondence of response button (left arrow vs. right arrow on the computer keyboard) to sound.

The experiment was run on the PsychoPy experiment-platform (Peirce, 2007, 2009).

Participants

Thirty-nine native French speaking participants (31 female, 35 right-handed) were run at Université Paris Descartes (average age 22.28, standard deviation 2.37).

Results

For each breathing position, each participant's data from the four staircases was submitted to a logistic regression to determine what point on the continuum corresponded to their perceptual boundary between /ada/ and /ana/ (the point at which they would hear the sound equally often as /ada/ and /ana/). These calculated boundaries were the dependent measure for the experiment. As this was a within-subjects design, the individual variability in perceptual boundaries (which are often highly variable between individuals) is not an issue here – each participant served as his/her own control.

As predicted, participants heard significantly more /ana/ when breathing through their nose than when breathing through their mouth, as determined by a paired t-test [$t(38) = 2.08$, $p = .044$, $d = .33$].

In the control experiment (categorizing /aga/ vs. /ada/), no such difference was found between breathing conditions [$t(38) = 1.36$, $p = .18$, $d = .21$]. However, the interaction between experiment and control versions did not reach significance. We believe this is due to a lack of power and are creating a new version of the experiment to address this issue.

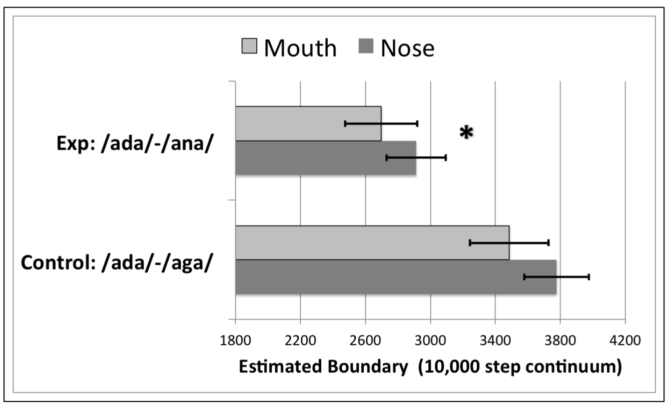

Figure 2: Results — Confidence Intervals are Shown

Discussion & Conclusion

Components of our motor systems, known as *forward models*, constantly predict the sensory effects of our actions — this prediction is *corollary discharge*. Corollary discharge serves a variety of crucial roles, including providing feedback for actions performed too quickly to use 'regular' sensory feedback and tagging self-produced sensations as such, thus preventing sensory confusion. It is this second role that allows corollary discharge to influence the concurrent perception of external sensations.

A recent group of theories (Skipper et al., 2006, 2007; Schwartz et al., 2010) have suggested that this function of corollary discharge may regularly be used to supplement perception in cases of perceptual uncertainty; generating a prediction, on the basis of one's own motor system, to guide the sensory processing. Forward models necessarily consult the current position of a person's articulators when generating corollary discharge, as sensory consequences are strongly dependent on the starting point of the effectors. This leads to the prediction that the position of one's own articulators should influence the perception of external speech sounds when those sounds are ambiguous and thus draw on the motor system's prediction abilities.

This experiment tested that prediction and has shown that the position of one's own articulators *does* influence the perception of the speech — even when the position of the articulators is adopted for a non-speech activity (breathing). These results support theories which argue for a role of the motor system (and corollary discharge) in speech perception and makes a unique contribution in showing that the static position of the articulators can have this effect even when their position is not intended to produce speech.

These results are relevant to the ongoing debate about embodied cognition — the degree to which the body and motor control systems are used in cognition. In the realm of semantic processing of language a similar debate is ongoing about the degree of motor involvement in the processing of the *meaning* of sentences. For example, the *Action Sentence Compatibility* effect demonstrates that movements of the arm are faster when a person reads a sentence implying arm movements, suggesting that the person's motor plan for arm movements was triggered by reading the sentence (Glenberg & Kaschak, 2002). The current experiment demonstrates a related example of embodied cognition, but at a 'lower', perceptual level of language processing.

Ongoing research is currently exploring the extent of this effect, examining how widespread (in terms of speech sounds) such effects are.

Acknowledgements

Funding for this project was provided by a United Arab Emirates University Research Start-Up Grant (Perceptual-Motor Linkages in Speech and Cognition - 31h060) to Mark Scott.

References

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009, April). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802.

Cornsweet, T. N. (1962, September). The Staircase-Method in Psychophysics. *The American Journal of Psychology*, *75*(3), 485–491.

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, *453*(7198), 1102–1106.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*(1), 3–28.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Goldstein, L. M., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In

Phonetics and phonology in language comprehension and production: Differences and similarities (pp. 159–207). Berlin: Mouton de Gruyter.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145.

Hickok, G., Houde, J., & Rong, F. (2011, February). Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron*, *69*(3), 407–22.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*.

Ito, T., Tiede, M., & Ostry, D. J. (2009, January). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1245–8.

Jones, J. A., & Munhall, K. G. (2002). *The role of auditory feedback during phonation: studies of Mandarin tone production* (Vol. 30) (No. 3).

Kawahara, H., Irino, T., & Morise, M. (2011). An interference-free representation of instantaneous frequency of periodic signals and its application to F0 extraction. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5420–5423).

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008* (pp. 3933–3936).

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Merckelbach, H., & Ven, V. V. D. (2001). Another White Christmas: fantasy proneness and reports of 'hallucinatory experiences' in undergraduate students. *Journal of Behavior Therapy and Experimental Psychiatry*, *32*, 137–144.

Peirce, J. W. (2007, May). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10.

Pickering, M. J., & Garrod, S. (2007, March). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–10.

Repp, B. H., & Knoblich, G. (2009). Performed or observed keyboard actions affect pianists' judgements of relative pitch. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*(11), 2156–2170.

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*(2-3), 429–435.

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2010, January). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 1–19.

Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, *11*(8), 349–355.

Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *Journal of the Acoustical Society of America Express Letters*, *133*(4), 286–293.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 250–285). Cambridge: Cambridge University Press.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007, October). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(10), 2387–99.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), 729–732.