Journal of Phonetics 56 (2016) 124–130

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

# A spatial gradient in phonetic recalibration by lipread speech

Mirjam Keetels*, Jeroen J. Stekelenburg, Jean Vroomen

*Department of Cognitive Neuroscience and Neuropsychology, Tilburg University, Tilburg, The Netherlands*

A R T I C L E I N F O

*Article history:*
Received 2 June 2015
Received in revised form
15 February 2016
Accepted 22 February 2016

*Keywords:*
Phonetic recalibration
Lip reading
Speech perception
Location specificity
Intersensory
Generalization

A B S T R A C T

Exposure to ambiguous speech combined with clear lipread speech can recalibrate auditory speech identification, a phenomenon known as phonetic recalibration (Bertelson, Vroomen, & De Gelder, 2003). Here, we examined whether phonetic recalibration is spatially specific. Participants were presented an ambiguous auditory sound halfway between /b/ and /d/ (A?) combined with lipread /b/ or /d/ at either the left or right ear/side, and were subsequently tested with auditory-only test sounds at either the same or the opposite ear/side. Phonetic recalibration was always strongest if test sounds were presented at the same ear/side than if they were presented at a different ear/side. Phonetic recalibration thus has a spatial gradient, showing that stimulus-specific and non-linguistic factors contribute to this phenomenon.



## 1. Introduction

Seeing a speaker can have a strong impact on the percept of the spoken message: it improves speech recognition in noise (Sumby & Pollack, 1954), and with incongruent audiovisual dubbings, lipread speech can even qualitatively alter the auditory percept (McGurk & MacDonald, 1976). More recent studies have demonstrated that lipread speech can also *recalibrate* auditory phoneme recognition in such a way that exposure to ambiguous speech combined with non-ambiguous lipreading can result in adjusted phoneme perception afterwards. Several studies have examined whether phonetic recalibration generalizes to other tokens, speakers, or phonetic contexts (Reinisch, Wozny, Mitterer, & Holt, 2014; Van der Zande, Jesse, & Cutler, 2014). One proposal is that recalibration is *speaker- or token-specific* because the effect is usually substantial if the test sound is the same and belongs to the adapted speaker, but minimal or non-existent if the test sound is different or likely belongs to another speaker. Here, we examined whether this lack of generalization to other tokens or speakers could originate from *non*-linguistic rather than linguistic reasons by asking whether recalibration is *spatially specific.* We report that phonetic recalibration is much smaller if the test sound comes from a different location than the adapted one. This spatial-specificity of phonetic recalibration highlights that a lack of generalization to other speakers, tokens, or phonetic contexts may also have *non*-linguistic roots.

To examine phonetic recalibration, we modified a procedure developed by Bertelson, Vroomen, and De Gelder (2003). In the typical paradigm, participants are exposed to an ambiguous speech sound intermediate between /aba/ and /ada/ (A?) dubbed onto a video of a face articulating either /aba/ or /ada/ (henceforth: A?Vb or A?Vd). During a subsequent auditory-only test phase, ambiguous speech sounds near the phoneme boundary are then identified as either /b/ or /d/. The typical result is that A? is *more* likely perceived as /b/ after exposure to A?Vb than after exposure to A?Vd. This learning effect occurs because, presumably, the conflict between sound and vision is reduced by a shift in the auditory phoneme boundary. No recalibration is observed if participants are exposed to perceptually similar, but non-ambiguous and audiovisual congruent tokens AdVd and AbVb. In this case, no learning effect is observed because there is no intersensory conflict between the heard and lipread token that evokes recalibration. Rather, these tokens evoke a contrastive effect (*less* /b/-responses after exposure to AbVb than after exposure to AdVd) indicative of selective speech adaptation or 'fatigue' of some linguistic feature detectors (Bertelson et al., 2003; Eimas & Corbit, 1973; Samuel,

* Corresponding author. Tel.: +31 13 4668187.
*E-mail address:* M.N.Keetels@TilburgUniversity.edu (M. Keetels).

0095-4470/$ - see front matter 
http://dx.doi.org/10.1016/j.wocn.2016.02.005

1986). Phonetic recalibration by lipread speech has now been replicated many times, also in other laboratories with other tokens and other phonemes. It is known that it builds-up fast (Vroomen & Baart, 2009; Vroomen, Van Linden, De Gelder, & Bertelson, 2007; Vroomen, Van Linden, Keetels, De Gelder, & Bertelson, 2004), and that it also can be induced by lexical information (Ganong, 1980; Kleinschmidt & Jaeger, 2015; Norris, McQueen, & Cutler, 2003; Samuel & Kraljic, 2009; Samuel & Lieblich, 2014; Van Linden & Vroomen, 2007). Recalibration by lipread speech probably involves early cortical auditory networks (i.e. regions within the anterior planum temporale, tangent to the posterior bank of Heschl's gyrus and sulcus) which shows that these areas respond differently to physically identical ambiguous sounds that are perceived differently (Kilian-Hutten, Valente, Vroomen, & Formisano, 2011; Kilian-Hutten, Vroomen, & Formisano, 2011).

Recent studies have focused on whether phonetic recalibration is specific to the exact phoneme heard during exposure, or whether it generalizes to a broader range of phonemes, contexts, or speakers. Van der Zande et al. (2014) examined whether phonetic recalibration by lipread speech generalizes across auditory test tokens of two different male speakers. Using similar procedures as Bertelson et al. (2003), these authors found large within-speaker effects, but recalibration was much smaller (though still significant) if the test stimuli came from a different speaker identity than the one in the adaptation phase. From a linguistic point of view, this decline makes sense because if one adapts to the voice or accent of a new speaker, this remapping might be less beneficial if applied to another speaker with a new accent. Furthermore, Reinisch et al. (2014) reported that phonetic recalibration by lipread speech is specific to the trained stimulus. In their study, recalibration was found for several different kinds of exposure stimuli (/aba/-/ada/; /ibi/-/idi/; /ama/-/ana/) but the effect never generalized to other test stimuli than the exposed ones. Reinisch et al. therefore concluded that phonetic recalibration is specific to the exact phoneme that has been adjusted during exposure.

In the present study, we examined whether this apparent lack of generalization across speakers and phonemes could be, at least partly, due to non-linguistic effects. It might be the case that recalibration is 'stimulus-specific' and that it thus is closely tied to the exact stimulus and context encountered during the exposure phase, *irrespective* of whether this context has linguistic meaning or not. A similar argument about the non-linguistic stimulus-specificity of lexical memory has recently been made by Pufahl and Samuel (2014). They reported that seemingly irrelevant information, such as an unattended background sound, is retained in memory and can facilitate subsequent speech perception. Here, we addressed a similar question by examining whether phonetic recalibration is *ear*- or *spatially-specific.* If so, one expects that if an ambiguous sound is adapted at, say location 'X', recalibration is stronger if the test sound is also presented at that specific location rather than another one. From a linguistic point of view, though, it should be noted that sound location is completely irrelevant because speech sounds do not change identity if a listener or speaker moves to another location. From a sensory-specific point of view, though, there are examples in the literature where visual learning (of e.g. global motion in random dots kinematograms) has been found to be extremely stimulus-, position, and even eye-specific (Crist, Kapadia, Westheimer, & Gilbert, 1997). Could it be, then, that phonetic recalibration is equally stimulus-specific?

A recent study by Keetels, Pecoraro, and Vroomen (2015) indeed hints in this direction. These authors reported that participants can be simultaneously recalibrated towards two different interpretations of the same sound if the exposure stimuli are presented in different ears. They used a procedure during which A? was presented in alternating fashion in the left and right ear via headphones while a concurrent left-right alternating video of a speaker was shown who articulated /b/ or /d/. Results showed that the A? was simultaneously adapted toward /b/ in the /b/-adapted ear and toward /d/ in the /d/-adapted ear, and that phonetic recalibration thus is specific for the exposed location.

In the present study, we examined whether this ear-specific recalibration effect was triggered by the rather idiosyncratic concurrent left-right exposure regime, or whether phonetic recalibration more generally carries information about where the adapted sound came from. In the latter case, one expects that exposure towards a single phoneme– thus *without* variation in sound location during exposure – nevertheless has a spatial gradient so that at test, effects are largest if test sounds are presented at the exact exposure location. To examine this, we modified the exposure-test paradigm used by Bertelson et al. (2003) so that participants were exposed to either A?Vb or A?Vd at one ear or side (left or right), and were subsequently tested with auditory-only sounds from either the same or a different ear or side. Non-ambiguous exposure stimuli (AbVb and AdVd), for which no recalibration was expected, served as baseline. In Experiment 1, sounds were presented via headphones, thus testing whether phonetic recalibration generalizes across *ears*. In Experiment 2, sounds were presented via two loudspeakers, so that they appeared to emanate from external space, testing whether phonetic recalibration generalizes across auditory *space*. We conceived three patterns of spatial generalization possible: (1) spatially-specific, (2) complete transfer, or (3) partial transfer in which phonetic recalibration is less strong – but still present – for test sounds at the opposite ear/side as exposure as compared to the same ear/side. Both experiments showed that there was partial transfer, and that phonetic recalibration thus always has a spatial ingredient.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Participants

Twenty students from Tilburg University participated and received course credits for their participation (8 male; 19 right-handed; 23.45 years average age). Participants reported normal hearing and normal or corrected-to-normal seeing. They were tested individually and were unaware of the purpose of the experiment. Written informed consent was obtained from each participant.