Journal of Phonetics 67 (2018) 65–77

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

The perception of anticipatory labial coarticulation by blind listeners in noise: A comparison with sighted listeners in audio-only, visual-only and audiovisual conditions

Véronique Delvaux [a,b,*], Kathy Huet [b], Myriam Piccaluga [b], Bernard Harmegnies [b]

[a] Fonds de la Recherche Scientifique – FNRS, 5, rue d'Egmont, 1000 Bruxelles, Belgium
[b] Institut de Recherche en Sciences et Technologies du Langage, Université de Mons, 20, place du Parc, 7000 Mons, Belgium

ARTICLE INFO

Article history:
Received 9 January 2017
Received in revised form 8 January 2018
Accepted 13 January 2018
Available online 6 February 2018

Keywords:
Audiovisual speech perception
Anticipatory coarticulation
Gating
Blind
French

ABSTRACT

This study investigates the time course of the perception of the /i-y/ contrast by French-speaking blind listeners using a gating paradigm. The performances of the blind listeners in discrimination and identification are compared with the range of performances exhibited by sighted perceivers when stimuli are presented auditorily, visually and audiovisually, whether in acoustically non degraded or in noisy conditions. Results provide evidence in favor of partial compensation for visual deprivation in speech perception. Blind listeners outperformed sighted participants in discriminating between auditorily-presented gated stimuli, particularly in noisy conditions. But this small advantage allowed them to compensate only partially for their inability to exploit visual information in order to process coarticulated speech as quickly and efficiently as sighted controls.

1. Introduction

1.1. Audiovisual speech perception

Speech perception is multimodal. Information from the auditory and visual modalities is integrated in speech processing, as evidenced by the classical McGurk effect (for a recent review, see Tiippana, 2014). Audiovisual integration exploits the time-varying shared properties between the acoustic and visual signals which result from their structural coupling within the talking individual (for a review of these properties, see Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). Even haptic information participates in speech processing when available, since manual tactile information relevant to recovering speech gestures has been shown to modulate auditory speech perception in untrained perceivers (e.g. Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer, 2008; Sato, Cavé, Ménard, & Brasseur, 2010).

Soon after birth, infants already seem to attend to both the auditory and visual speech inputs, although multisensory integration sophisticates throughout the first year of age (Lewkowicz & Pons, 2013). Audiovisual speech perception has been shown to play a role in language development, including in speech production skills (e.g. Desjardins, Rogers, & Werker, 1997), so that an impairment in audiovisual speech perception and/or in the underlying integration process, may be associated with a number of developmental communication disorders (Guiraud et al., 2012; Leybaert et al., 2014; Woodhouse, Hickson, & Dodd, 2009).

Neuroimaging and neurophysiological literature strongly support multimodal speech perception in adult speakers (for a review, see Alsius, MacDonald, & Munhall, 2013), but the nature (i.e., amodal vs. modality-specific) of the information processing code, and the stage of the integration during processing (i.e., early, late or both) remain debated (Dias, Cook, & Rosenblum, 2016; Peelle & Sommers, 2015; Rosenblum, 2008; Woodhouse et al., 2009).

Whereas there is a number of cross-predictabilities which makes the integration of the information from the auditory and visual streams possible (Schwartz & Savariaux, 2014), auditory and visual modalities typically convey complementary information, so that audiovisual integration, when possible, improves speech perception. Indeed, visual information

* Corresponding author at: IRSTL, Université de Mons, 20, Place du Parc, 7000 Mons, Belgium. Tel.: +32 65373140.

E-mail address: Veronique.Delvaux@umons.ac.be (V. Delvaux).

https://doi.org/10.1016/j.wocn.2018.01.001
0095-4470/

enhances overall speech intelligibility, especially in adverse conditions, e.g. in noisy environments (Bernstein, Auer, & Moore, 2004; Grant & Seitz, 2000; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Schwartz, Berthommier, & Savariaux, 2004; Sumby & Pollack, 1954), in nonnative listeners (Navarra & Soto-Faraco, 2007; Wang, Behne, & Jiang, 2009), or in listeners with impaired hearing (Bergeson, Pisoni, & Davis, 2005; Winn, Rhone, Chatterjee, & Idsardi, 2013). Visual information is used in speech perception when local auditory information is ambiguous, as evidenced in studies on the use of visual contextual cues to speaker identity (Mitchel, Gerfen, & Weiss, 2016), and visually-guided perceptual recalibration (Bertelson, Vroomen, & de Gelder, 2003; Reinisch & Mitterer, 2016). When congruent visual information is available, listeners also perform better and/or more rapidly in other, related tasks, such as perceptual learning (Hazan, Sennema, Faulkner, & Ortega-Llebaria, 2006; Wayne & Johnsrude, 2012) or shadowing in noisy conditions (Scarbel, Beautemps, Schwartz, & Sato, 2014). It is noteworthy that audiovisual speech perception is prone to individual variation, which may partly reflect the effects of socio-cultural (including linguistic) factors as well as subject-dependent factors among which intelligence, age, gender, hearing (dis)ability, working memory and modality dominance (Schwartz, 2010; Sekiyama, Burnham, Tam, & Erdener, 2003; Sekiyama, Soshi, & Sakamoto, 2014; Stevenson, Zemtsov, & Wallace, 2012; Woodhouse et al., 2009).

Peelle and Sommers (2015) attribute the visual enhancement effect to two main types of information gathered by seeing a speaker's mouth: (i) the supra-segmental information (associated with temporal amplitude envelope) provided by the mouth opening and closing gestures, which may improve the listeners' predictions about the upcoming speech signal, thus facilitate sound onset detection (see also Baart, 2016) and increase perception accuracy over the entire speech signal; and (ii) the segmental information provided by specific articulator position, which may assist word recognition by constraining lexical competition. In fact, head motion and movements of the different regions of the face – including eyebrows, cheeks, and chin – have been shown to carry valuable information for speech perception, mostly in the prosodic domain (Chen & Massaro, 2008; Granström, House, & Lundeberg, 1999; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Yehia, Kuratate, & Vatikiotis-Bateson, 2002). In terms of segmental perception, visual cues are especially informative on place of articulation in consonants (Jesse & Massaro, 2010), and rounding in vowels (Traunmüller & Öhrström, 2007). Jesse and Massaro (2010) further showed that visual and auditory speech do not only differ in what featural information they convey, but also in when they do so over the speech signal. In other words, visual and auditory information are not temporally aligned in speech.

In French, the determination of the rounding feature in vowels is heavily based on the visual channel for sighted speakers, either in static configurations (i.e. for sustained vowels: Robert-Ribes, Schwartz, Lallouache, & Escudier, 1998) or in dynamic configurations (i.e. in anticipatory labial coarticulation, in $V_{[-\mathrm{rounded}]}(C)C_{[-\mathrm{labial}]}V_{[+\mathrm{rounded}]}$ sequences: Abry, Lallouache, & Cathiard, 1996; Cathiard, 1994; Roy, 2012; Troille, Cathiard, & Abry, 2010). As pointed out by Schwartz and Savariaux (2014), anticipatory labial coarticulation provides an interesting case study of the temporal desynchronization between the visual and the audio streams of information in connected speech: while the visual rounding gesture may anticipate the vowel by up to 200–300 ms (e.g. Abry et al., 1996), and typically results in the rounded vowel being detected earlier in audiovisual speech (Cathiard, 1994), the visual enhancement effect may be mediated or even annihilated by the specific timing of coordinations between the involved articulatory gestures (Troille et al., 2010).

1.2. Speech perception in visually-impaired people

By definition, most of visual information is not available for speech perception in visually-impaired people. However, a few studies have shown that even when visual acuity is severely reduced, providing audiovisual stimuli still enhances speech perception in noise in younger as well as in older visually-impaired adults (Gagné & Wittich, 2010; Hickson, Hollins, Lind, Worrall, & Lovie-Kitchin, 2004).

When vision loss is total, there is substantial evidence of cross-modal plasticity, with visual cortical activity observed for various tasks involving sound processing (Amedi, Floel, Knecht, Zohary, & Cohen, 2004; Bedny, Pascual-Leone, Dodell-Feder, Fedorenko, & Saxe, 2011; Burton, Diamond, & McDermott, 2003; Gougoux, Zatorre, Lassonde, Voss, & Lepore, 2005; Poirier et al., 2006; Röder, Stock, Bien, Neville, & Rösler, 2002), at least when blindness is congenital or early acquired (Bedny, Pascual-Leone, Dravida, & Saxe, 2012). This neuronal reorganization is considered as part of the compensatory mechanisms which allow effective perceptual processing of auditory information despite the sensory deprivation (for a review, see Occelli, Spence, & Zampini, 2013).

Behavioral studies have documented enhanced performance of blind listeners over sighted controls in a wide range of auditory tasks, such as auditory spatial tuning (Fieger, Röder, Teder-Sälejärvi, Hillyard, & Neville, 2006; Röder et al., 1999) echolocation (Dufour, Després, & Candas, 2005; Teng, Puri, & Whitney, 2012), processing of simple sounds like tones (Niemeyer & Starlinger, 1981; Röder, Rösler, Hennighausen, & Näcker, 1996), pitch detection (Gougoux et al., 2004; Wan, Wood, Reutens, & Wilson, 2010), absolute pitch (Hamilton, Pascual-Leone, & Schlaug, 2004), and identification of voices (Braun, 2012; Bull, Rathborn, & Clifford, 1983). It is noteworthy, however, that not all studies and/or auditory-related tasks resulted in blind listeners outperforming sighted listeners (Gougoux et al., 2009; Gunzburger, Bresser, & Ter Keurs, 1987; Starlinger & Niemeyer, 1981; Wan et al., 2010; Winograd, Kerr, & Spence, 1984; Zwiers, Van Opstal, & Cruysberg, 2001). It remains unclear whether this outcome is speaker- or task-related, or if some publication bias is at play here.

Concerning speech sounds specifically, there is a large body of evidence demonstrating improved processing of (synthetic or time-compressed) fast speech in blind people (Dietrich, Hertrich, & Ackermann, 2013; Gordon-Salant & Friedman, 2011; Trouvain, 2007), and scarcer evidence of enhanced intelligibility of speech material (words and sentences) in a noisy environment (Chen, Liu, & Chen, 2014;