Journal of Phonetics 65 (2017) 94–109

Contents lists available at ScienceDirect

Journal of Phonetics

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

Research Article

Car-talk: Location-specific speech production and perception

Jennifer Hay[a,b,*], Ryan Podlubny[a,b], Katie Drager[c], Megan McAuliffe[a,d]

[a] New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
[b] Department of Linguistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
[c] Department of Linguistics, University of Hawai'i at Mānoa, 561 Moore Hall, 1890 East-West Road, Honolulu, HI 96822, USA
[d] Department of Communication Disorders, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

ARTICLE INFO

Article history:
Received 11 October 2016
Received in revised form 21 June 2017
Accepted 28 June 2017
Available online 29 July 2017

Keywords:
Lombard effect
Speech perception
Socioindexicality
Speech in noise
Contextual probability

ABSTRACT

Some locations are probabilistically associated with certain types of speech. Most speech that is encountered in a car, for example, will have Lombard-like characteristics as a result of having been produced in the context of car noise. We examine the hypothesis that the association between cars and Lombard speech will trigger Lombard-like speaking and listening behaviour when a person is physically present in a car, even in the absence of noise. Production and perception tasks were conducted, in noise and in quiet, in both a lab and a parked car. The results show that speech produced in a quiet car resembles speech produced in the context of car noise. Additionally, we find tentative evidence indicating that listeners in a quiet car adjust their vowel boundaries in a manner that suggests that they interpreted the speech as though it were Lombard speech.

1. Introduction

It has long been known that individuals display remarkable flexibility in speech production. Some of this variation appears to be highly agentive, as speakers tailor their pronunciations to their identity-related goals (Coupland, 2007; Eckert, 2000; Podesva & Callier, 2015). However, some variation also appears to be highly automatic, based on acoustic or non-acoustic primes in the environment (Delvaux & Soquet, 2007; Sanchez, Hay, & Nilson, 2015). Additionally, it has been shown that high levels of flexibility also exist in perception, with perceptual boundaries between phonemes affected by listener beliefs about speaker identity (Johnson, Strand, & D'Imperio, 1999; Drager, 2011), attitudes (Walker, Hay, Drager, & Sanchez, in press), and social primes not directly related to the speaker (Hay & Drager, 2010). Taken together, this body of work demonstrates how experience-driven probabilistic expectations influence both speech production and perception.

Such results are frequently accounted for in a usage-based account of production and perception, which posits that individuals store acoustically and contextually rich memories of their past experiences with speech, and that these memories affect subsequent production and perception patterns. Activating contextual primes or identity goals will activate relevant subsets of these memories, thus biasing production and perception patterns in the appropriate direction (see e.g. Foulkes & Docherty, 2006; Foulkes & Hay, 2015; Pierrehumbert, 2006, 2016). Additionally, such models make the untested prediction that "language would also have a strong association with location, suggesting that changes in environment should cause a shift in which phonetic variants are produced and perceived" (Hay & Drager, 2007, p. 98). Indeed, this prediction would follow from any model in which contextual probabilities, derived from prior experience, contribute to the likelihood of producing or perceiving a sound a certain way (see, e,g. Nielsen & Wilson, 2008; Norris & McQueen, 2008).

In the current paper, we set out to test this prediction explicitly. We ask: do individuals shift their vowel production and perception in contextually-specific ways that are consistent with their prior experience? Specifically, we look at car-specific speech production and perception. Most speech encountered in a car was produced in the context of car noise. Therefore, speech that was produced and encountered in a car will have had characteristics of Lombard speech (i.e., speech produced in noise).

* Corresponding author at: New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch, New Zealand.

E-mail addresses: jen.hay@canterbury.ac.nz (J. Hay), ryan.podlubny@pg.canterbury.ac.nz (R. Podlubny), kdrager@hawaii.edu (K. Drager), megan.mcauliffe@canterbury.ac.nz (M. McAuliffe).

http://dx.doi.org/10.1016/j.wocn.2017.06.005
0095-4470/

We tested whether being physically present in a non-noisy car induces Lombard effects in production that are similar to those observed when talking over car noise. We then tested whether – when the sounds are played in car noise and/or when the participants are physically present in the car – listeners adapt their perception of sounds accordingly. We hypothesized that being physically present in a car would induce effects on an individual's production and perception that were consistent with car noise, even when no car noise was present.

2. Background

2.1. The Lombard effect

The Lombard effect was first identified as an increase in speech intensity when talkers compete with noise (Lombard, 1911). The effect was later recognised to include increases in F0 and word duration, as well as a levelling of spectral tilt, as a result of increases in energy in the higher frequency bands (Brumm & Zollinger, 2011; Cooke & Lu, 2010; Draegert, 1951; Junqua, 1993; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988; Zhao & Jurafsky, 2009; Tartter, Gomes, & Litwin, 1993). The Lombard effect has been documented in a wide range of background noises, including car noise (Jung, 2012). Formants are also affected by environmental factors. Van Summers et al. (1988) observed F1 frequencies rising in noise, though the size of the effect was variable and not always present across talkers. Similarly, Pisoni, Bernacki, Nusbaum, and Yuchtman (1985) observed a lowering of F2 within sufficiently noisy conditions. However, the presence of this effect was likewise inconsistent.

It is well understood that inter-speaker differences that arise given physiology and/or speech techniques account for some of this variation. Additionally, variability in the presence or strength of reported Lombard effects across projects may stem from variation in speakers' characteristics (Egan, 1972) or differences in experimental design, such as the task required (Lane & Tranel, 1971) or type of background noise used (Cooke & Lu, 2010; Parikh & Loizou, 2005; Stowe & Golob, 2013).

Previous work on the perception of Lombard speech focuses on how it is perceived more accurately than non-Lombard speech in noise (Dreher & O'Neill, 1957; Lu & Cooke, 2008), and demonstrates that synthesized Lombard speech in noise is perceived with accuracy comparable to naturally produced Lombard speech (Raitio, Suni, Vainio, & Alku, 2013). However, it has not tested for changes in listening strategy or perceived vowel boundaries when listeners are listening in noise. If speakers systematically shift their realizations of sounds – shifting, for example, the first and second formants of a vowel in response to a noisy environment – then we hypothesize that listeners could potentially use this information when interpreting an acoustic signal in noise, shifting their perceptual boundaries to be in line with the speech they have previously encountered in noisy environments.

2.2. Memory for background noise

Due to the Lombard effects outlined above, our experience of speech produced in background noise systematically differs from our experience of speech which was not produced in noise. This might lead us to predict that past experience of speech produced in noise will affect the perception of future speech produced in a similarly noisy context. But what role does ambient noise actually play in a model in which acoustically detailed memories are stored?

Listeners are highly proficient at accurately attending to a voice signal in the face of potentially interfering noise (see e.g. Bronkhorst (2015) and Mattys, Davis, Bradlow, & Scott (2012) for reviews relating to this process). But once the speech in noise has been accurately processed, what are the implications for subsequent storage? One can imagine at least two possibilities. The first is that the acoustic details of the background noise are stored as part and parcel of the acoustic trace. The second is that the listener attends to the speech and filters out the background noise as unnecessary for storage. In either case, the listener may store the exemplar complete with information about the fact that it was produced in noise (even if the details of that noise are not stored).

Some recent work supports the first possibility, demonstrating that background noise is stored in a way that affects subsequent perception, at least in the short term (Creel, Aslin, & Tanenhaus, 2012; Cooper, Brouwer, & Bradlow, 2015; Pufahl & Samuel, 2014). Creel et al. (2012) trained participants on novel words, presented in either noise or quiet. They show that if the test stimulus is presented in the same noise-condition as the training stimulus, this improves participant performance on the test. Likewise, Pufahl and Samuel (2014) conducted a series of experiments investigating the role of background noises such as car horns or dogs barking. They found that minor changes to the background noise (such as changing to a bark from a small dog rather than a bigger dog) interfered with word recall. Based on the results of a series of six experiments, they argue that: "Seemingly irrelevant information, such as an unattended background sound, is retained in memory and can facilitate subsequent speech perception" (Pufahl & Samuel, 2014, p. 1). In a follow-up experiment to Pufahl and Samuel (2014), Strori (2016) found results that were consistent with the suggestion that background noise can be encoded in representation, and further argues that: "unlike indexical effects, sound specificity effects are fragile and conditional. Listeners seem to be able to encode details of sounds co-occurring with speech in their memory representations, but only in certain occasions" (Strori, 2016, p. 146). Along these same lines, Cooper et al. (2015) report results from a speeded classification paradigm, in which irrelevant variation in background noise slowed classification of speech according to speaker gender or speaker identity. They also report results from a continuous recognition memory paradigm, in which participants were explicitly asked whether a recognized word was new, or whether it was an old word, presented with an old or new background noise. They found that the background sound affected word recognition, but only when there was some spectral overlap between the noise and the word.

Much of this research is still new and additional work is required to understand the exact circumstances in which the reported effects arise. However, the results from these experiments suggest that auditory memories of background noise may be closely linked with the associated speech memories. Thus, it seems likely that, in at least some contexts, Lombard speech may be stored together with the background noise, or with the abstracted information that the speech was produced in noise.