Available online at www.sciencedirect.com
SciVerse ScienceDirect
Computer Speech and Language 28 (2014) 580–597
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?☆

Maëva Garnier*, Nathalie Henrich

Department of Speech and Cognition, GIPSA-Lab (UMR 5216: CNRS, INPG, University Stendhal, UJF), Grenoble, France

Received 14 October 2012; received in revised form 26 July 2013; accepted 29 July 2013
Available online 15 August 2013

Abstract

What makes speech produced in the presence of noise (Lombard speech) more intelligible than conversational speech produced in quiet conditions? This study investigates the hypothesis that speakers modify their speech in the presence of noise in such a way that acoustic contrasts between their speech and the background noise are enhanced, which would improve speech audibility.

Ten French speakers were recorded while playing an interactive game first in quiet condition, then in two types of noisy conditions with different spectral characteristics: a broadband noise (BB) and a cocktail-party noise (CKTL), both played over loudspeakers at 86 dB SPL.

Similarly to (Lu and Cooke, 2009b), our results suggest no systematic "active" adaptation of the whole speech spectrum or vocal intensity to the spectral characteristics of the ambient noise. Regardless of the type of noise, the gender or the type of speech segment, the primary strategy was to speak louder in noise, with a greater adaptation in BB noise and an emphasis on vowels rather than any type of consonants.

Active strategies were evidenced, but were subtle and of second order to the primary strategy of speaking louder: for each gender, fundamental frequency (f_0) and first formant frequency (F1) were modified in cocktail-party noise in a way that optimized the release in energetic masking induced by this type of noise. Furthermore, speakers showed two additional modifications as compared to shouted speech, which therefore cannot be interpreted in terms of vocal effort only: they enhanced the modulation of their speech in f_0 and vocal intensity and they boosted their speech spectrum specifically around 3 kHz, in the region of maximum ear sensitivity associated with the actor's or singer's formant.

Keywords: Lombard speech; Noise; Production; Speech audibility; Auditory detection; Segregation; Energetic masking

1. Introduction

Noise exposure triggers an adaptation in speech production, commonly referred to as the Lombard effect. When communicating in noisy environments, speakers commonly increase vocal intensity and fundamental frequency (f_0) as compared to communicating in quiet environments (Castellanos et al., 1996; Junqua, 1993; Van Summers et al., 1988). Speech produced in noise (also called Lombard speech) is also characterized by a higher first-formant frequency of vowels (F1), boosted energy above 2 kHz and increased vowel/consonant (*V*/*C*) ratio in both vocal intensity and

☆ This paper has been recommended for acceptance by 'Dr. Martin Cooke'.

* Corresponding author. Tel.: +33 4 76 57 50 61.
E-mail address: maeva.garnier@gipsa-lab.grenoble-inp.fr (M. Garnier).

0885-2308/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.csl.2013.07.005

duration (Boril and Pollak, 2005; Castellanos et al., 1996; Egan, 1972; Junqua, 1993; Kadiri, 1998; Mokbel, 1992; Stanton et al., 1988; Van Summers et al., 1988).

Lombard speech has been shown to be more intelligible than conversational speech produced in quiet condition (Dreher and O'Neill, 1957; Lu and Cooke, 2008; Pittman and Wiley, 2001; Van Summers et al., 1988). It remains unclear which speech modifications contribute to this gain in intelligibility and which aspects of intelligibility are improved by this speech adaptation (phoneme recognition, speech audibility, utterance parsing, etc.).

In this article, we focus on whether and how speakers may try to improve their audibility in noise, i.e. the detection and perception of speech information by their interlocutor within the background noise. Before envisaging and suggesting some possible strategies, let us first review the current knowledge on the mechanisms and factors that influence speech perception in noise.

First, it is known that the audibility of a sound is considerably degraded when it is heard simultaneously with a competing noise or sound stream that contains energy in the same critical frequency bands. The energetic-masking effect increases with increased spectral overlap and decreased signal-to-noise (SNR) ratio between the target sound and the masker (Hornsby and Ricketts, 2001; French and Steinberg, 1947). In the case of speech, multi-talker noise degrades the perception of vowels more than consonants, whereas white Gaussian noise has the opposite effect (Junqua, 1993). Similarly, speech is more degraded by a competing speech produced by a speaker of the same gender (Brungart, 2001).

Auditory fusion is another perceptual phenomenon that occurs when two or more sound streams are heard at the same time. The concurrent streams are interpreted as coming from the same source when they present similar acoustic characteristics (such as the average intensity, pitch and timbre, but also in the temporal modulation of these parameters). They are segregated from each other when acoustic contrasts exceed a given threshold (Darwin et al., 2003).

Both phenomena of energetic masking and auditory fusion underlie the “cocktail-party effect”, i.e. the difficulty of following a voice and understanding what is said within a multitude of other competing voices (Arons, 1992). Psychoacoustic research showed how the segregation of a target speech from another competing speech is particularly difficult when both voices are similar in spectral content and fundamental frequency (f_0) (Assmann and Summerfield, 1990), first-formant frequency (F1) (Darwin et al., 2003), when the target voice is not modulated in f_0 (Marin and McAdams, 1991), and when it is compressed in amplitude (Hornsby and Ricketts, 2001).

What could speakers then do to improve their speech audibility and segregation in noise?

(1) Speakers may try to decrease the amount of energetic masking and enhance acoustic contrasts by increasing the global vocal intensity of their speech, and more specifically the spectral energy in frequency regions where the background noise presents maximum energy (boosting strategies).
(2) They may try to shift the spectral energy, or at least important phonetic cues coded in frequency (f_0, formants), to spectral bands where the background noise presents minimum energy (bypass strategies).
(3) They may try to increase the temporal modulation of their speech in f_0 and vocal intensity (modulation strategies).

Evidence of boosting strategies was provided by Mokbel (1992) and Junqua et al. (1998). Mokbel (1992) demonstrated that a speaker enhances his speech energy more in the frequency band where noise was concentrated. In a single-talker experiment comparing speech adaptation to broadband noises filtered by different band-pass filters, Junqua et al. (1998) showed that, at constant masker level, the increase of vocal intensity varies with noise spectral tilt. Bypass strategies were observed in recent studies (Lu and Cooke, 2008, 2009a,b) that showed how the center of gravity (CoG) of the speech spectrum increases in frequency when speaking in low-pass noises (multi-babble noise, driving noise or low-pass filtered broadband noise). However, they did not observe such bypass strategies in high-pass filtered noises (Lu and Cooke, 2009a,b), in which the speech CoG and additional spectral cues (f_0, F1) were not shifted down but were still shifted to higher frequencies. No study has specifically explored the use of modulation strategies in noise. Nevertheless, enhanced intonation contours and wider f_0 ranges in noise have been reported by several authors (Boril and Pollak, 2005; Garnier et al., 2006, 2010; Welby, 2006).

In line with these previous studies, this study aims at examining whether speakers adopt boosting, bypass or modulation strategies in noise to adapt to the spectral characteristics of the background noise and enhance acoustic contrasts between their speech and that noise. Two different types of noise frequently encountered in ecological situations were chosen here to test speech adaptation to varying energetic masking: a broadband noise (with equally distributed energy below 10 kHz) and a cocktail-party noise (with concentrated energy below 1 kHz). Similarly to previous studies, evidence of boosting, bypass or modulation strategies was searched in the global adaptation of speech