Effect of level and type of noise on focus related prosody

Martti Vainio[1], *Antti Suni*[1], *Anja Arnhold*[1,3], *Tuomo Raitio*[2], *Henri Seijo*[2], *Juhani Järvikivi*[4],Daniel Aalto[1,2], and *Paavo Alku*[2]

[1]Institute of Behavioural Sciences (SigMe Group), University of Helsinki, Finland
[2]Department of Signal Processing and Acoustics, Aalto University, Finland
[3]Department of Cognitive Linguistics, Goethe-University Frankfurt am Main, Germany
[4]Language Acquisition and Language Processing Lab, NTNU, Norway
martti.vainio@helsinki.fi

Speakers automatically raise their voice when forced to speak in environmental noise or when the normal feedback mechanism is disturbed. Raising one's voice consists of various physiological means that have different consequences on the phonetic realisation of speech. Typically the speakers' f_0 is higher – probably as a consequence of added sub-glottal pressure – and the mode of vocal fold vibration is more pressed in order to decrease the slope of the glottal voice-source spectrum. The adaptation of speech to noise in order to increase the signal-to-noise ratio is called the Lombard effect or Lombard reflex to illustrate its involuntary nature [1]. However, the knowledge regarding Lombard speech is fairly general in nature and not very much is known about how the reflex influences prosodic changes that are due to specific communicative needs such as signalling prosodic focus.

We recorded 21 speakers producing utterances with different information focus conditions in three types of noise with four noise levels. The purpose of the study was to see whether speakers vary their means of producing prosodic focus as a function of both noise level and type. The analysed utterances were replies to three types of questions designed to elicit either a broad focus or two types of narrow focus in simple three word utterances. The typical prosodic patterns for the three focus conditions are well-known for Finnish, which allows us to compare Lombard speech to normal speech in a controlled manner. The three types of noise were: white noise, babble noise, and a 1 kHz low-pass noise. The noises were scaled for equal loudness on three different levels corresponding to approximately 60, 70 and 80 dB(A) sound pressure levels.

We analyzed the produced utterances with regard to f_0, *duration*, *voice source features*, *formants* and *intensity*. Only pitch related features are presented here. The pitch contours were analyzed in terms of three different points per word: the pitch maximum (peak) and the minima left and right of it (valleys). Thus, there are nine potential values for each utterance. The contours clearly follow the typical shapes associated with different focus conditions in Finnish [2, 3]; i.e, the narrowly focused word has a higher peak and post-focal words have lower peaks but are not altogether deaccented. The verbs also have a rising-falling shape, but with a markedly lower magnitude [4]. For further analyses the f_0 values were transformed to semitones (re 100 Hz).

The f_0 *expansion* was calculated from the nine f_0 values per sentence (also in semitones) by adding the absolute differences. This can be expressed in terms of an integral based on the Bounded Variation (BV) norm:

$$\text{Expansion}(f_0) = \int_{T_{beg}}^{T_{end}} \left| \frac{df_0(s)}{dt} \right| ds \qquad (1)$$

where $f_0(t)$ is the fundamental frequency at a given time point and T_{beg} and T_{end} are the beginning and end times of the utterance.

Statistical analyses with linear mixed-effects models show that with regard to mean f_0 the noise levels differ significantly. The low-pass noise has a significantly lower mean f_0. As can be expected from the results of previous studies, the different focus types are also different from each other: i.e, the f_0 is generally lower when the narrow focus occurs on the first word and higher when it occurs on the last word. There is also a significant low-pass-noise:noiselevel3 interaction showing that the f_0 level is increased less in high level low-pass noise. With regard to f_0 expansion the results show that the contours are significantly influenced by the focus type as well as noise levels. The low-pass noise, however, does not differ from babble noise, but the contours are again more expanded in white noise (t = 3.64). This is also shown in the white-noise:level interactions. There are no significant gender differences.

In summary, the analysis of the data shows that, regardless of the increase in f_0, the typical intonation contours are still produced. Also, and as can be expected, the f_0 is raised as a function of noise level. Moreover, the intonation contour is expanded as the f_0 gets higher; the louder the noise, the more expanded the contour is. In addition the noise type affects the contour in different ways and and there are interesting level-type interactions. The typical utterance-final creaky voice is also not as prevalent in Lombard speech as it is normally and disappears altogether in severely noisy conditions.

1. References

[1] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, vol. 37, no. 101-119, p. 25, 1911.

[2] M. Vainio and J. Järvikivi, "Focus in production: Tonal shape, intensity and word order," *The Journal of the Acoustical Society of America*, vol. 121, no. 2, pp. EL55–EL61, 2007. [Online]. Available: http://link.aip.org/link/?JAS/121/EL55/1

[3] A. Arnhold, "Multiple prosodic parameters signaling information structure: Parallel focus marking in finnish," in *Proceedings of the International Congress of Phonetic Sciences (ICPhS), Hong Kong*, 2011.

[4] A. Arnhold, M. Vainio, A. Suni, and J. Järvikivi, "Intonation of Finnish verbs," in *Speech Prosody 2010-Fifth International Conference*, 2010.