

Intonational Differences in Lombard Speech: Looking Beyond F_0 Range

Pauline Welby

Institut de la Communication Parlée, CNRS UMR 5009

Institut National Polytechnique de Grenoble, Université Stendhal, Grenoble, France

welby@icp.inpg.fr

Abstract

Previous studies on speech in noise (Lombard speech) have generally reported an increase in fundamental frequency (F_0). This study examines three other potential intonational differences: choice of intonation pattern, tonal scaling, and tonal alignment. Seven French speakers read a corpus of short paragraphs, in quiet and in 80 dB white noise. Four of the speakers increased F_0 range across the target accentual phrases in noise. Six speakers upscaled individual tones; there was great inter-speaker variability in tonal scaling, in contrast with an earlier study on Dutch. No influence of noise on intonation pattern type was found. In particular, there was no tendency to produce more “early rises” in noise, even though these rises are cues to word segmentation. Producing an early rise (thus a LHLH or LHH pattern) may not add to the salience of the commonly produced LH pattern. In addition, no difference in tonal alignment was found, in contrast to the findings of an earlier study. This null result may be due to paradigm differences between the two experiments.

Further work on intonational differences in Lombard speech should concentrate on aspects beyond F_0 range or global averages, including those that may be language-specific.

1. Introduction

1.1 Lombard effect

Many conversations take place in some sort of noise – children playing, cars passing, wind whistling, to name a few. But speakers adapt – they speak louder, alter the duration of speech segments (generally increasing the duration of vowels as compared to consonants), and they raise their fundamental frequency (F_0) (see [8, 10, 16] and references therein on these and other changes). Speech changes in noise are collectively referred to as the Lombard effect or reflex.

At least some of these changes render speech more intelligible, thus helping listeners and more effectively transmitting the message. In one perception experiment, tokens produced in 90 dB of masking noise and tokens produced in quiet were mixed with noise at various signal-to-noise ratios and presented to participants. Listeners more accurately identified tokens produced in noise than those produced in quiet [16]. There are, however, limits to the intelligibility gains of Lombard speech; for example, shouted speech is less intelligible ([12], see also [8]).

A number of studies have examined the effect of noise on F_0 , an effect first noted by Lombard himself. These studies generally report increased F_0 and considerable interspeaker variability. One study of 10 speakers found a large increase in F_0 (“between 82 and 106 Hz”) for male speakers and a smaller increase for female speakers speaking in 85 dB white noise [8]. In another study of two male speakers, F_0 differences were found from quiet to three different noise conditions [16]. For one speaker, this change was quite small (< 5 Hz); the second speaker raised his F_0 almost 17 Hz from quiet to 80 dB noise.

Very few studies, however, have examined the influence of noise on details of intonational structure. Indeed, studies reporting F_0 changes have typically used lists of isolated,

mostly monosyllabic words, which are not ideal for examining intonational structure. One exception is a study that modeled the scaling of F_0 turning points (peaks and valleys) “of low internal variability” in Dutch sentences read by 15 speakers in quiet and in noise [14].¹ Speakers raised F_0 for all turning points in noise. The authors concluded that the “raising function” could be described with a single model for all turning points, except the final low. One goal of the current study was to examine the scaling of turning points in French.

1.2 Some details about French intonation

All accounts of French intonation agree that the utterance is divided into smaller units, called Rhythmic Phrase [4], Accentual Phrase (AP) [7], etc. by various researchers. This AP is typically characterized by an F_0 rise on the last syllable of a phrase that is not utterance-final (late or final rise) and an optional early (initial) rise near the beginning of the phrase. Examples from the current study are given in Figure 1, with Autosegmental Metrical (AM) transcriptions. In this framework, intonation patterns are treated as a series of L(ow)

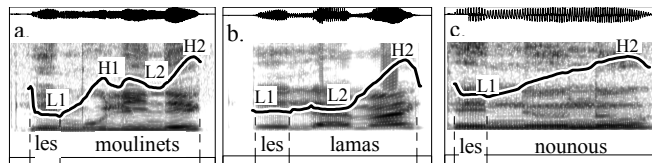


Figure 1: APs with (a) an early and a late rise (LHLH), (b) a late rise (LLH), (c) a rise from an early L to a late H (LH).

and H(igh) tones, with intermediate values determined by interpolation [13]. Another possible pattern, often found in short APs, is LH, a rise from L1 to H2 (without H1 or L2).

As Figures 1a and 1c illustrate, the beginning of the early rise (L1) is aligned to the beginning of the first content word of the AP [17, 19], often forming an “elbow” in the F_0 curve when there is a preceding function word. This elbow is often found even in the absence of a following rise.

An early rise can be used as a cue to word segmentation [18, 19]. Listeners interpret an ambiguous sequence like [me.la.mõ.din] as a four-syllable pseudo-content word (*mélamondine*) when there is an early rise starting at the first syllable [me], but as a function word followed by a three-syllable pseudo-content word (*mes lamondines*) ‘my lamondines’ when the rise starts at the second syllable [la]. Furthermore, they can use even an F_0 elbow (not followed by a rise) to find the beginning of a content word.

¹ The authors themselves do not consider their comparison to involve Lombard speech: “We assume that a considerable portion of the pitch change obtains because speakers deliberately raise their pitch; this is the phenomenon we intend to describe. We recognize, however, that some portion of the changes will also occur as a by-product of an increase in vocal effort (the “Lombard reflex”...),” but the motivations for this claim are not described.

1.3 Recent findings on intonational changes in Lombard speech in French

A recent study examined articulatory and intonational changes by a single native speaker of French in noisy speaking conditions [2, 20]. Three intonational differences between speech in quiet and speech in noise were found. Speech was recorded in a quiet condition, in the presence of cocktail party noise, and in the presence of white noise. The noise was presented at 85 dB over loudspeakers positioned two meters from the participant and was later subtracted from the signal by a channel-estimation algorithm.

Early rises: We had hypothesized that speakers might produce more early rises in noise than in quiet, since those rises should make speech easier to segment and thus more intelligible. The results were in the direction of the hypothesis – the speaker produced more APs with early rises in the noise conditions. Yet, the results must be interpreted with caution: first, they are for only a single speaker and second, the APs in the study were longer in the noisy speech conditions than in the quiet control condition. This duration difference is not unexpected given reports in the literature, but presents a confound, since longer APs are known to favor the realization of early rises [17, 19].

AP-internal tonal scaling: The speaker significantly raised her F_0 range from the quiet control condition to the noise conditions. But she also showed an unexpected tonal scaling pattern: in APs with the two-rise LHLH pattern, the fall from the peak of the early rise (H1) to the beginning of the late rise (L2) was much smaller in noise than in quiet. In fact, L2 often had an F_0 value around that of the preceding H1, producing a plateau, as in Figure 2b.



Figure 2. Schema of tonal scaling difference in LHLH patterns. (a) clear “dip” from H1 to L2, (b) plateau from H1 to L2.

The tonal scaling pattern found in Figure 2b is not uncommon in French; patterns like Figures 2a and b represent end points of a continuum, not a dichotomy.² Yet patterns like Figure 2b are more common in shorter APs, in which speakers undershoot the L2 target under time pressure (see [19], pp. 77–80). The finding in [20] is surprising because the speaker produced this pattern (LHplateauLH) in *longer* APs.

A final finding of the study was that early rises produced in noise did not always start at the function word/ content word boundary, but surprisingly, sometimes started at the beginning of the *function word*. This unusual alignment may be due to a lack of auditory feedback in noise [5].

1.4 Goals of the current study

The goal of the current study was to examine the influence of noise on three aspects of intonational structure raised by the earlier study: the speaker’s choice of intonation pattern, the scaling of individual tones, and the alignment of these tones.

2. Methods

2.1. Materials

A set of seven short paragraphs was designed, each containing two to five target words. There were 11 two-syllable target words and 11 three-syllable target words. Target words were

chosen with the help of the online database LEXIQUE [11]. Targets contained only CV syllables with sonorant consonants (either nasals or the lateral /l/). Using only sonorants minimized segmental perturbations of the F_0 curve. Each target word appeared sentence-initially, preceded by an article (*le, la, les* or *un*). The resulting phrase was followed by relative clause beginning with a relative pronoun (*qui, que, qu’il*), followed by the verb phrase of the main clause. An example paragraph, with targets underlined is given in (1).

(1) Il y eut une scène chaotique à la crèche cet après-midi. Un mulot qui s’était réfugié au fond de la cantine avait fait peur aux petits. Ils s’étaient tous mis sur leurs chaises en hurlant. Les moulinets que Daniel était en train de faire avec les bras avaient fait tomber le bocal à poissons. Il y avait de l’eau et du verre partout. Heureusement, Charlotte pensa à mettre le pauvre petit poisson rouge dans un verre d’eau. Les nounous qui étaient venues chercher les enfants secouaient la tête en regardant la scène.

‘There was a chaotic scene at the daycare center this afternoon. A field mouse that had taken refuge in the back of the lunchroom had scared the little ones. They had all climbed up on the chairs screaming. The windmills that Daniel was making with his arms had knocked over the fishbowl. There was water and glass everywhere. Luckily Charlotte had thought to put the poor little goldfish in a glass of water. The nannies that had come to pick up the children shook their heads as they watched the scene.’

Each participant was seated across a small table from the experimenter and instructed to “read the paragraphs as if you were reading them to the person across from you.”

2.2. Participants

Participants were seven female native speakers of Hexagonal French. Speaker 7 was the speaker in the experiment in [20].

2.3. Procedures

Participants first familiarized themselves with the paragraphs by reading them silently. They then read the corpus aloud four times, first in quiet at a self-selected normal rate and at a fast rate, then in noise at a normal rate and at a fast rate. Each speaker read the paragraphs in a different randomized order.

White noise (generated using Cool Edit Pro 2.0 software, with a one-second ramp) was presented to participants over headphones at 80 dB. Speakers were recorded at 22.05 kHz using a Shure SM10A headworn microphone and a Marantz PMD670 digital recorder and the data transferred to computer. Recording was done in a sound attenuated chamber at the Institut de la Communication Parlée.

Two rates were included only as a mechanism for obtaining speech in quiet and in noise with comparable durations for target APs to avoid the confound discussed for the earlier study ([20]). I had hoped that the tendency of speakers to slow down in reading the paragraphs in noise would be somewhat mitigated by the fact that the normal rate noise condition immediately followed the fast rate quiet condition (i.e., that there would be some carryover from the fast rate).

The sound files were segmented and each utterance saved as a separate file. F_0 curves and spectrograms were created using Praat speech analysis software [3]. Word boundaries were tagged for each target sequence, using waveforms and spectrograms to guide the segmentation. These boundaries were generally easy to identify, since the target regions contained nasals or liquids followed by vowels, and there were thus clear changes in intensity. The beginning and end of each utterance were also tagged. Praat scripts were written to semi-automate the segmentation and labelling process. These

² This distribution argues against accounts in which these two contours (one with a clear dip from H1 to L2, and one with a plateau) form two distinct entries in an inventory of patterns (e.g., [15]).

prompted the user to click on the location of a desired tag, automatically inserted boundaries and tags, and saved the results to a Praat TextGrid file. A line-fitting procedure was used to automatically detect Lows. (See [17, 19] for details; for the scripts see: <http://www.icp.inpg.fr/~welby/praat.html>.)

3. Results

3.1. Rate manipulation

As discussed above, the goal of the rate manipulation was to minimize or eliminate duration differences between the normal rate quiet condition and the normal rate noise condition. As Figure 3 shows, this manipulation was largely unsuccessful. Only one speaker, Speaker 3, had AP durations in those two conditions that were not significantly different ($t(21) = .110$; $p = .914$). A second speaker, Speaker 6, had longer AP durations in the normal rate noise condition than in the normal rate quiet condition ($t(21) = -4.379$; $p < .001$). The remaining five speakers had shorter AP durations in the normal rate noise condition (all p values $< .01$, except for Speaker 4, $p < .05$).

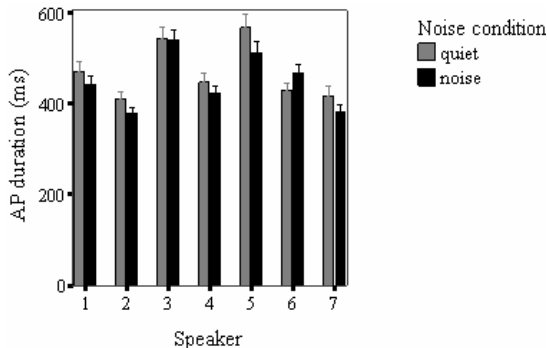


Figure 3: AP durations for the two normal rate conditions. Bars show standard error of the mean.

3.2. Intonation patterns found

Contrary to the predictions, none of the speakers showed a significant increase in the production of early rises from the quiet condition to the noise condition.

Almost 90% of three-syllable target words were produced with an early rise, most with the LHLH pattern, some with the LHH pattern (in which L2 is not realized).

For the two-syllable target words, there was more variety in intonation patterns. Only 15% of APs in the normal rate quiet and normal rate noise conditions were produced with the LHH or LHLH patterns, while 84% were produced with either the LH or the LLH pattern. Three of the seven speakers seemed to prefer the LH pattern, but three speakers produced many LLH patterns. Speaker 7 produced many LLH patterns, but also a number of LHLH patterns.

3.3. Tonal alignment

I visually inspected the tonal alignment of the beginning of the early rise (L1) in the normal rate quiet condition and the normal rate noise condition. Short APs with the LH pattern often have no elbow and the rise can begin at the very

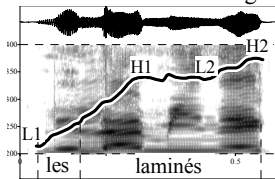


Figure 4: The AP *les laminés* ‘the laminated ones’ realized with a LHLH pattern. The early rise begins at the beginning of the function word *les*. (from normal rate noise condition, Speaker 6).

beginning of the AP. Longer APs with the LHLH pattern, however, typically have L1 elbows. Unlike the previous study, there was no evidence for these APs that L1 was less precisely aligned in the noise condition; most had L1 elbows at the function word-content word boundary. There were, however, occasional cases of LHLH APs in which L1 was realized at the beginning of the function word, as in Figure 4.

3.4. Tonal scaling

F_0 range across the target AP (the difference between H2 and L1) was examined; four speakers raised their F_0 range from the quiet to the noise condition, as illustrated in Figure 5.

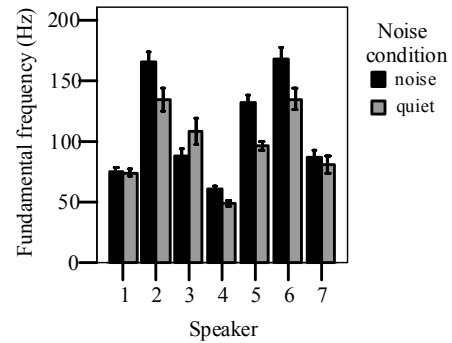


Figure 5: Mean F_0 range in target APs by noise condition. Bars show standard error of the mean.

In comparing the scaling of individual tones, it is not possible to compare intonation patterns of different types, since details of tonal scaling may vary across patterns (e.g., L2 tends to be lower in LLH than in LHLH [19]). The comparisons here are therefore confined to LHLH, the most commonly produced pattern. One-factor ANOVAs were performed to examine whether noise condition influenced the scaling of each of the four tones (L1, H1, L2, H2). The results are shown in Table 1.³

Table 1: F_0 values for the four tones in the LHLH pattern by condition. All significant differences are marked with a ‘*’.

S	L1		H1		L2		H2	
	quiet	noise	Quiet	noise	quiet	noise	quiet	noise
1	197	198	253	257	235	249	265	281
2	203	217*	257	272	251	290*	316	393*
3	193	199	270	260	214	216	279	274
5	208	227*	304	337*	260	292*	309	355*
6	216	231	305	331*	272	296*	334	373
7	179	186*	248	242	201	221*	255	265

As Table 1 shows, only Speakers 2 and 7 (the speaker in [20]) show something approaching the ‘LHplateauLH’ pattern found in [20]. In noise, these speakers raised L2 without raising H1, resulting a smaller dip from H1 to L2.

4. Discussion and Conclusions

Of the parameters examined, the results were most fruitful for scaling. For all but one speaker, there were significant differences in the scaling of individual tones between the quiet and the noise conditions, although there was great inter-speaker variability in which tone or tones were upscaled. This variability is unsurprising, since work on French has shown inter-speaker scaling differences (see [6, 21]). The finding of differences in the scaling of individual tones is of particular

³ Speaker 4’s data were excluded since she produced too few (three) LHLH patterns in the noise condition.

The differences between the results of the current study, in which seven French speakers showed great variability in upscaling of individual tones, and those in [14], in which 15 Dutch speakers consistently upscaled all tones, may be language-specific. We know that tonal scaling in clear speech varies from language-to-language (e.g., French vs. Dutch and Greek [1]). So too may tonal scaling in Lombard speech. Since the exact F_0 shape of the AP does not generally signal a meaning difference in French (at least for unambiguous APs), speakers may not be under pressure to preserve all scaling details.

The current study did not uncover any evidence of less precise L1 alignment in noise, even in Speaker 7 (the speaker in [20]). This may be due to differences in the experimental paradigm. The earlier study used 85 dB noise presented over loud speakers (placing not only the speaker, but also the experimenter/listener in a noisy environment), while the current study used 80 dB noise presented over headphones. Tonal alignment may only be perturbed past a certain intensity threshold or with a certain degree of vocal effort.

A number of authors have stressed the influence of the task in Lombard speech studies (e.g., [9]). In particular, increasing communicative load (e.g., by using speaker-listener pairs) may lead to enhancements of some aspects of the effect.

Studies of intonational changes in Lombard speech have both theoretical implications and potential applications. For example, results on F_0 range and scaling show which aspects of relative tone height are preserved under pressure in a given language and are thus relevant to intonational phonology. They also help us to understand the extent to which these changes may be language-specific. A better understanding of these changes will allow us to model them in developing speech technologies. For example, if intonational changes increase intelligibility in noise (an open question), these changes could be incorporated into speech synthesis systems designed to be used in noisy environments.

[1] Arvaniti, A., Ladd, D. R., Mennen, I., 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *J Phon.* 26, 3–25.

- [2] Bailly, L., 2005. Étude articuloire de la parole produite en environnement bruyant. Master's thesis. Laboratoire d'Acoustique Musicale, Université de Paris 6.
http://www.www.icp.inpg.fr/~lbailly/RapportFinal_LucieBailly.pdf
- [3] Boersma, P., Weenink, D., 2005. PRAAT: Doing phonetics by computer (version 4.3.28) [Computer program], <http://www.praat.org>
- [4] Di Cristo, A. (1998). Intonation in French. In D. Hirst and A. Di Cristo (eds.), *Intonation Systems: A survey of twenty languages*, pp. 195–218. Cambridge: CUP.
- [5] Elliott, L., Niemoeller, A., 1970. The role of hearing in controlling voice fundamental frequency. *Internatl. Audiology* 9, 47–52.
- [6] Fougeron, C., Jun, S.-A., 1998. Rate effects on French intonation: Prosodic organization and phonetic Realization. *J. Phon.* 26, 45–69.
- [7] Jun, S.-A., Fougeron, C., 2000. A phonological model of French intonation. In *Intonation: Analysis, modelling and technology*, pp. 209–242. A. Botinis (ed.). Dordrecht/ Boston: Kluwer.
- [8] Junqua, J.-C., 1993. The Lombard reflex and its role on human listeners and automatic speech recognizers. *JASA* 93, 510–524.
- [9] Lane, H., Bernard, T., 1971. The Lombard sign and the role of hearing in speech. *J. Speech and Hearing Research* 14, 677–709.
- [10] Lombard, É., 1911. Le signe de l'élévation de la voix. *Annales des Maladies de l'oreille et du larynx*, 37, 101–119.
- [11] New, B., Pallier, C., Ferrand, L., Matos, R., 2001. Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'Année Psychologique* 101, 447–462. Available from <http://www.lexique.org>
- [12] Pickett, J. M., 1956. Effects of vocal force on the intelligibility of speech sounds. *JASA* 28, 902–905.
- [13] Pierrehumbert, J. B., 1980. *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT.
- [14] Shriberg, E., Ladd, D. R., Terken, J., Stolcke, A., 1996. Modeling pitch range variation within and across speakers: Predicting F0 targets when “speaking up.” *ICSLP 4*, Philadelphia, pp. 1–4 (proceedings addendum). Corrected version available at <http://www.asel.udel.edu/icslp/cdrom/vol2/553/a553.pdf>
- [15] Vaissière, J., 1977. Une procédure de segmentation automatique de la parole en mots prosodiques en français. Actes des VIIèmes Journées d'Étude sur la Parole, pp. 103–114. Nancy.
- [16] Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., Stokes, M.A., 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *JASA* 84, 917–928.
- [17] Welby, P., in press. French intonational structure: Evidence from tonal alignment. *J. Phon.*
- [18] Welby, P., 2003. French intonational rises and their role in speech segmentation. *Eurospeech* 8, pp. 2125–2128. Geneva.
- [19] Welby, P., 2003. *The slaying of Lady Mondegreen, being a study of French tonal alignment and association and their role in speech segmentation*. PhD dissertation. The Ohio State University. <http://www.ling.ohio-state.edu/publications/dissertations>
- [20] Welby, P., Bailly, L., Garnier, M., Dohen, M., Lævenbruck, H., submitted. Pitch changes in noisy conditions: Data from French. *Proceedings of the Conference Making Europe More Attractive for Researchers*. Pisa/Livorno.
- [21] Welby, P., Lævenbruck, H., to appear. Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics*. Special issue on Autosegmental-metrical approaches to intonation in Europe: Tonal targets and anchors, M. D'Imperio (ed).

I thank Hélène Løevenbruck for advice in designing the study and analyzing the data and for valuable discussions, Denis Beauteemps for helpful comments, and Denis Beauteemps, Xavier Pelorson and Coriandre Vilain for technical assistance. I am also grateful to the participants. This research was supported by a Marie Curie International Fellowship within the 6th European Community Framework Programme.