STRATEGIES OF VOCAL ADAPTATION TO BACKGROUND NOISE OF DYSPHONIC AND CONTROL SCHOOLTEACHERS

Amelia PETTIROSSI[1,2], Nicolas AUDIBERT[1], Lise CREVIER-BUCHMAN[1,3]

[1]Laboratoire de Phonétique et Phonologie, UMR7018 Université Sorbonne Nouvelle/CNRS, Paris, France
[2]Imagerie & Cerveau, UMR1253 INSERM/Université de Tours, Tours, France
[3]Hôpital Foch : Service de Laryngologie Phoniatrie, Université Paris Saclay, Suresnes, France

amelia.pettirossi@univ-tours.fr, nicolas.audibert@sorbonne-nouvelle.fr, lise.buchman1@gmail.com

ABSTRACT

This paper aims at observing strategies of vocal adaptation to background noise in a population of 37 control and 24 mild dysphonic female schoolteachers, recorded in a quiet "neutral" reading context and then by picturing reading in front of a noisy classroom.

Beyond the expected consequences of the Lombard effect, vocal coping strategies are set-up in a noisy environment by all speakers, with steeper spectral slopes and longer reading times. Observed strategies may be interpreted as a way to be better perceived by listeners.

Comparison of Long-Term Average Spectrum shapes between conditions and groups suggest that control speakers implement more efficient coping strategies with a greater low-frequency reinforcement that may be interpreted as a better vocal projection. On the other hand, this adaptability seems to be limited by organic capabilities related to the laryngeal pathophysiological conditions in dysphonic speakers.

Keywords: coping strategies, schoolteachers, dysphonia, quiet vs. noisy context, adaptability

1. INTRODUCTION

Functional dysphonia is due to a bad vocal gesture leading to a misuse or an overuse of the vocal folds, the consequence of which is an alteration of the voice quality and a discomfort in the oral communication process [1]. As voice is schoolteachers (ST) main "working tool", they are considered as voice professionals. However, French schoolteachers receive very little (if any) voice coaching, which results in them being highly affected by dysphonia. In fact, a literature review highlights that dysphonia is significantly more present in schoolteachers than in the rest of the general population and this is particularly true for women, who are a majority in this profession since they constituted 83% of the entrants in 2000 [2].

This high prevalence of dysphonia is, among others, the consequence of their evolution in a noisy environment. The average ambient noise in a classroom is about 72 dB [3], for comparison, a level of 80 dB corresponds to heavy road traffic. This constant background noise then forces STs to increase their intensity to be properly perceived by pupils: this phenomenon is called "Lombard effect" [4]. As a result, STs work in a constantly noisy environment, which has an effect on the vocal effort required to be heard. The "ideal" average intensity for conversational speech between a speaker and a listener in a quiet room is 50 dB at a distance of 1m, the ideal intensity then increases by 3 dB per 10 dB of background noise [5]. In a classroom averaging 72 dB, ST have to speak at an average of 9 dB higher than their conversational intensity during their working day. Even though the Lombard effect is an automatic environmental accommodation phenomenon also shown in non-human primates [6], it necessarily generates vocal forcing [2,4]. In fact, the increase of the intensity is correlated to the increase of the fundamental frequency [7,8]. Thus, rising the f0 leads to more microtraumas related to the contact of the vocal folds.

Furthermore, we can say that there is a vocal "overdose" among schoolteachers since they give 6 to 7 hours of class per day, 5 to 6 days a week, which results in much too short resting periods for the vocal folds [9].

We wish to investigate potential vocal adaptation strategies to background noise in a population of dysphonic and control female schoolteachers. Although we expect to observe some adaptability in both groups, we assume that this will be less pronounced in the dysphonic one.

2. METHODOLOGY

2.1. Participants, corpus and recording sessions

Before the beginning of our data collection, the following experimental protocol was evaluated and validated by an ethics committee specialized in behavioral and health research, the "*Comité d'Ethique pour les Recherches Comportementales et En Santé*" (CERCES) of Paris Descartes University.

61 female schoolteachers signed a consent form and a "right to voice" form before being recorded under controlled conditions to collect audio data.

They were all native French speakers, actively working in a public or private school.

Our analyses focus on two readings of the French version of "The north wind and the sun". The first is read in a quiet and "neutral" condition and the second one by picturing reading in front of a "noisy" classroom.

2.2. Perceptual evaluation of vocal pathology

Voice quality was assessed using the GRBAS scale [10, 11]. This perceptual evaluation scale assesses the grade of dysphonia (G), the roughness (R), the breathiness (B), the asthenia (A) and the vocal strain (S) by rating these vocal parameters from 0 (unimpaired) to 3 (severely impaired). Our evaluation is performed on a speech sample by a phoniatrist and a speech-therapist and reveals 24 mild dysphonic speakers (two G2 and 22 G1) and 37 speakers rated G0 considered as the "control" group.

2.3 Acoustic measures

We focus on analyses related to Long-Term Average Spectrum (LTAS). The pitch-corrected Long-Term Average Spectrum (pitch-corrected LTAS) is chosen as an alternative to the "classic" LTAS because of its greater robustness to intra-speaker variability [12]. The pitch-corrected LTAS and spectral slopes are computed from all voiced parts of the signal for frequencies below 10 kHz, with a bandwidth of 50 Hz.

3. RESULTS

First of all, when comparing the average spectral slopes and total reading times between control and dysphonic speakers, we observe limited differences between the two groups (Table 1).

Spectral slopes are on average steeper for both types of reading even if this phenomenon is more noticeable in the dysphonics compared to the controls. This difference between groups is significant in the "noisy" context. Similarly, the average total reading time is slightly longer for dysphonic speakers in the "noisy" condition, but this difference is not significant, and the standard deviation is also larger for this group.

The comparison between contexts, all speakers pooled, allow us to highlight a pattern of accommodation common to both groups in the "noisy" condition. Indeed, the spectral slopes are significantly lower in the "noisy" context than in the "neutral" one (Table 2).

At the same time, our data reveal significantly longer total reading times, still all speakers combined, when STs are reading in the "noisy" condition compared to the "neutral" environment (Table 2).

Spectral slope	Control	Dysphonic	Independent t-test
"Quiet" context	-24.47 (2.23)	-25.70 (2.92)	t(59)=1.831 p=0.07
"Noisy" context	-19.21 (2.86)	-20.96 (3.29)	t(59)=2.176 p=0.03
Total duration	**Control**	**Dysphonic**	**Independent t-test**
"Quiet" context	42.18 (3.51)	42.18 (4.68)	t(59)=0 p=0.50
"Noisy" context	45.38 (4.32)	46.63 (7.18)	t(59)=-0.84 p=0.20

Table 1: Spectral slopes (dB) and total reading time (sec) compared between control vs. dysphonic speakers in "quiet" and "noisy" context. Mean values are presented with standard deviation in parentheses.

Spectral slope	"Quiet" context	"Noisy" context	Paired t-test
All speakers pooled	-24.95 (2.59)	-19.90 (3.15)	t(60)= -16.12; p<0.0001
Total duration	**"Quiet" context**	**"Noisy" context**	**Paired t-test**
All speakers pooled	42.18 (4.01)	45.88 (5.65)	t(60)= -8.1; p<0.0001

Table 2: Spectral slopes (dB) and total reading time (sec) compared between "quiet" vs. "noisy" context, all speakers pooled. Mean values are presented with standard deviation in parentheses.

These results suggest that there is a common behaviour between dysphonics and controls even if dysphonics seems to be slightly more impacted by the "noisy" environment as we can see by comparing the LTAS shapes between conditions and groups.

The visual inspection of graphical representations of the average pitch-corrected LTAS of our two reading contexts confirm a common pattern between dysphonics and controls (Figure 1). In both cases we can see an increase of energy in the low frequencies from 1 to 3 kHz when reading in front of a "noisy" class, with a larger difference for control speakers. On the other hand, we can observe a difference in the high frequencies, since there is an increase of energy between 5 and 7 kHz for dysphonic speakers in both conditions, with a larger difference between groups of speakers in the “noisy” context.

Finally, we present the average difference between the “neutral” and “noisy” conditions after matching the values per speaker, compared between groups (Figure 2).

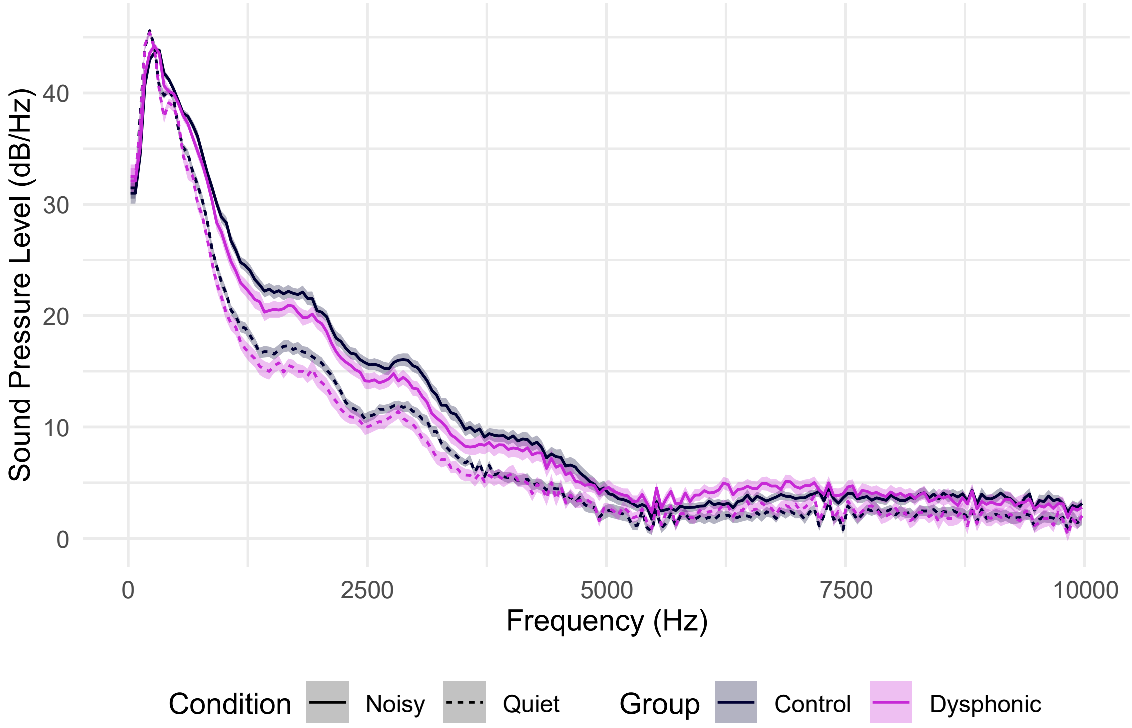

Figure 1: Average pitch-corrected LTAS with standard error of the control group (blue line) and the dysphonic group (purple line) during the reading in “neutral” condition (dotted line) and “noisy” condition (solid line)

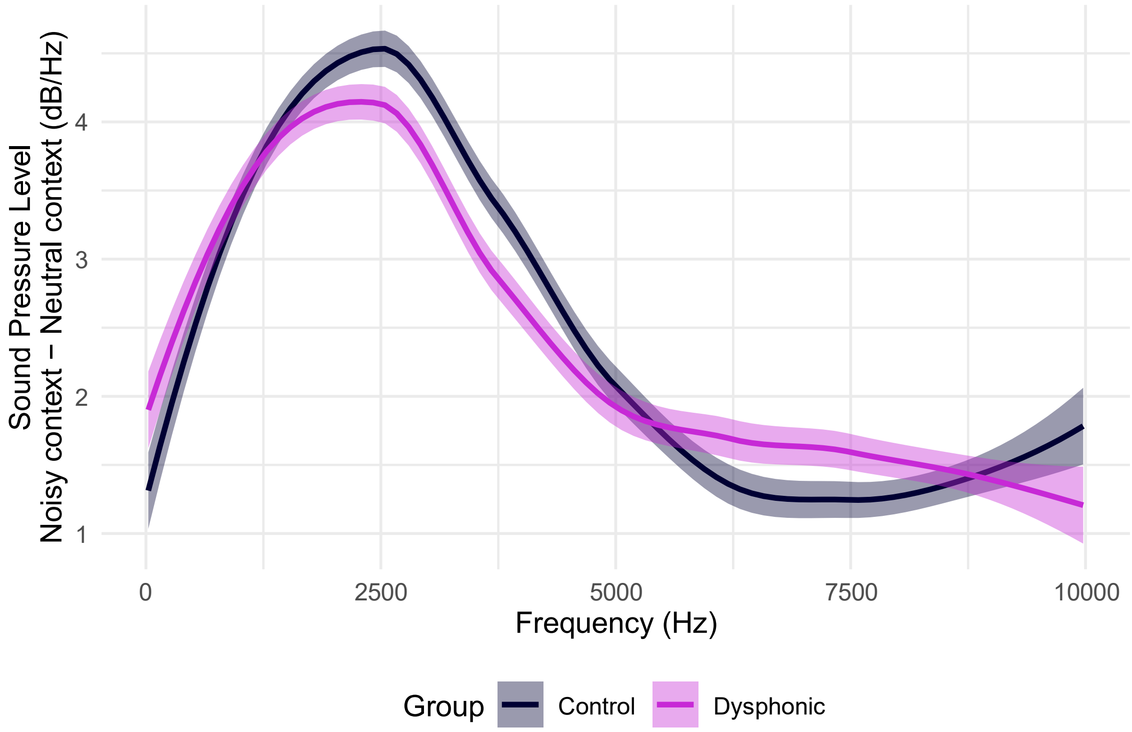

Figure 2: Mean difference in sound pressure level (dB/Hz) with standard error in the “neutral” vs. “noisy” condition smoothed by a local “loess” regression, for the dysphonic group (purple line) and the control group (blue line)

The curves presented in this figure correspond to the difference in acoustic energy in all frequency ranges when switching from one condition to another. In order to allow an easier visualization of the general trends for wider frequency ranges beyond local fluctuations, plotted values are smoothed by means of a “loess” local regression. This representation confirms a greater reinforcement of the energy in the low frequencies for controls. We also observe the previously discussed greater increase in high frequency energy for dysphonic schoolteachers in the “noisy” condition compared to the “neutral” one.

4. DISCUSSION AND PERSPECTIVES

Beyond the expected consequences of the Lombard effect [4], research looking at speaker’s ability to enhance acoustic contrasts between their voice and background noise concluded that there is an expected increase in intensity but also an amplification of the vocal spectrum around 3 kHz [13]. This increase in energy around 3 kHz is not observable in “quiet” context speech, nor in shouted speech [14]. The description of this energetically enhanced area perfectly matches what we observe for the schoolteachers analysed in our study. This phenomenon could therefore be the result of background noise compensation and not only of the expected increase in intensity and spectral modifications present in shouted speech. Furthermore, this reinforcement is stronger for the control group and it could be linked to the “actor's formant” which is a phenomenon similar to the “singer's formant”. The singer’s formant leads, when produced, to a better evaluation of voice quality [15]. While the literature clearly establishes the existence of the singer's formant for male voices, it is sometimes not observed for women [16]. Still, some studies conclude that there is an increase in energy between 4 and 5 kHz, which would then correspond to the actor's formant for female voices [17]. Although moderate, the increase in energy around 3 kHz observed in our data could be related to this phenomenon. Our results indeed suggest the use by female schoolteachers of a reinforcement of this frequency range in order to allow for more effective communication, especially in a noisy environment requiring an increased vocal efficiency. The difference observed between the dysphonic and control groups also suggests that the dysphonic speakers would have more difficulty reinforcing this frequency range to project their voice efficiently.

The observation of an increase of energy between 5 and 7 kHz for dysphonic speakers in both conditions is consistent with the literature since a high frequency peak can be considered as a signature of dysphonic speech [18]. Based on our data, this “acoustic signature” of dysphonia would be more salient when the vocal effort increases.

These vocal compensation strategies may be interpreted within the frame of Lindblom’s hypo- and hyper-articulation theory [19] as a response to communicative demands, in order to compensate for a noisy classroom environment or vocal impairment which would impair listener’s perception without such adaptation.

The principle of plasticity (and thus of hyper-articulation) aiming at making large gestures to be better perceived by the listener, may explain the observed longer total reading times and the increased energy between 2 and 3.5 kHz. Moreover, the fact that this phenomenon is more visible for control speakers can be interpreted as the consequence of better vocal projection capacities in this group whereas this adaptability seems to be limited by organic capabilities related to the laryngeal pathophysiological conditions in dysphonic speakers.

Finally, observation of our data leads us to question what seems to be a compensation mechanism to a noisy environment. Extending these analyses to reading tasks in multiple settings with different controlled intensity of background noise would provide a better understanding of the extent to which the negative factors of background noise and dysphonia may accumulate and require greater compensation to ensure a successful communication.

Following the covid-19 pandemic and the mandatory wearing of masks in schools, it would also be interesting to add this constraint to our future experimentations. Indeed, several studies have already shown a link between the wearing of protective masks and a decrease in intelligibility in classrooms [20, 21] or an alteration in the student-teacher relationship [22]. Future work will investigate to what extent such alterations could cumulate with communication impairment due to dysphonia.

5. CONCLUSION

We can therefore conclude that all the female schoolteachers use a common strategy of vocal adaptation to better project their voice in order to improve the contrast between their voice and the background noise of the classroom. On the other hand, we note a greater adaptability for the control group compared to the dysphonic group.

6. ACKNOWLEDGEMENTS

This work has been partially supported by the "Investissements d'Avenir" program ANR-10-

LABX-0083 (Labex EFL) and by ANR GEPETO (ANR-19-CE28-0018).

6. REFERENCES

[1] Crevier-Buchman, L., Tessier, C., Sauvignet, A., Arpin, S. B., & Monfrais-Pfauwadel, M.-C. (2005). Diagnostic d'une dysphonie non organique de l'adulte. Revue de Laryngologie Otologie Rhinologie, 126, 353-360.
[2] INSERM: Expertise Collective. (2006). La voix : Ses troubles chez les enseignants. Paris: INSERM, Institut national de la santé et de la recherche médicale.
[3] Shield, B., & Dockrell, J. E. (2004). External and internal noise surveys of London primary schools. The Journal of the Acoustical Society of America, 115(2), 730-738.
[4] Dejonckere, P. H., & Pépin, F. (1983). Etude de l'effet Lombard par la mesure du niveau sonore équivalent. Folia Phoniatrica et Logopaedica, 35(6), 310-315.
[5] van Heusden, E., Plomp, R., & Pols, L. C. W. (1979). Effect of ambient noise on the vocal output and the preferred listening level of conversational speech. Applied Acoustics, 12(1), 31-43.
[6] Ruch, H., Zürcher, Y., & Burkart, J. M. (2018). The function and mechanism of vocal accommodation in humans and other primates: Vocal accommodation in human and other primates. Biological Reviews, 93(2), 996-1013.
[7] Alku, P., Vintturi, J., & Vilkman, E. (2002). Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation. Speech Communication, 38(3-4), 321-334.
[8] Zee, E. (1977). Duration and intensity as correlates of F0. Journal of Phonetics, 6(3), 213-220.
[9] Titze, I. R. (1999). Toward occupational safety criteria for vocalization. Logopedics Phoniatrics Vocology, 24(2), 49-54.
[10] Hirano, Minoru. (1981). Clinical examination of voice. Wien; New York: Springer-Verlag.
[11] Hirano, M. (1989). Objective evaluation of the human voice: Clinical aspects. Folia Phoniatrica et Logopaedica, 41(2-3), 89-144.
[12] Pettirossi, A., Audibert, N., & Crevier-Buchman, L. (2017). Le spectre moyen à long-terme corrigé en F0 (pitch-corrected LTAS) : Une mesure robuste de la voix dysphonique sur des énoncés de nature variable. Présenté à 7èmes Journées de Phonétique Clinique, Paris, France.
[13] Garnier, M., & Henrich, N. (2014). Speaking in noise : How does the Lombard effect improve acoustic contrasts between speech and ambient noise? Computer Speech & Language, 28(2), 580-597.
[14] Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. The Journal of the Acoustical Society of America, 115(1), 362-378.
[15] Leino, T., Laukkanen, A.-M., & Radolf, V. (2011). Formation of the Actor's/Speaker's Formant : A Study Applying Spectrum Analysis and Computer Modeling. Journal of Voice, 25(2), 150-158.
[16] Master, S., De Biase, N. G., & Madureira, S. (2012). What About the "Actor's Formant" in Actresses' Voices? Journal of Voice, 26(3), 117-122.
[17] Tayal, S., Stone, S., & Birkholz, P. (2017). Towards the measurement of the actor's formant in female voices. Studientexte zur Sprachkommunikation: (p. 286-293). Présenté à Konferenz Elektronische Sprachsignalverarbeitung, Saarbrücken, Deutschland.
[18] Kitzing, P., & Åkerlund, L. E. I. F. (1993). Long-time average spectrograms of dysphonic voices before and after therapy. Folia Phoniatrica et Logopaedica, 45(2), 53-61.
[19] Lindblom, B. (1990). Explaining Phonetic Variation : A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Éds.), Speech Production and Speech Modelling (p. 403-439). Dordrecht: Springer Netherlands.
[20] Bottalico, P., Murgia, S., Puglisi, G. E., Astolfi, A., & Kirk, K. I. (2020). Effect of masks on speech intelligibility in auralized classrooms. The Journal of the Acoustical Society of America, 148(5), 2878-2884.
[21] Caniato, M., Marzi, A., & Gasparella, A. (2021). How much COVID-19 face protections influence speech intelligibility in classrooms?. Applied Acoustics, 178, 108051.
[22] Fykaris, I., Matiaki, V., & Geroulidou, D. C. (2021). The Mask of Personal Protection as a Cause of Depersonification of the Learners in Formal Education. Journal of Education, Society and Behavioural Science, 34(1), 30-41.