Is the perception of voice quality language-dependant? A comparison of French and Italian listeners and dysphonic speakers

Alain Ghio[1], *Frédérique Weisz*[2], *Giovanna Baracca*[3], *Giovanna Cantarella*[3], *Danièle Robert*[1&2], *Virginie Woisard*[1&4], *Franco Fussi*[5], *Antoine Giovanni*[1&2]

[1] LPL, Laboratoire Parole et Langage, CNRS, Aix-Marseille University, France
[2] Service ORL, CHU de la Timone, Marseille, France
[3] Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico ORL Dept, Milano, Italy
[4] Service ORL, Unité voix et déglutition, CHU Rangueil-Larrey, Toulouse, France
[5] Centro Audiologico Foniatrico Azienda USL Ravenna, Italy

alain.ghio@lpl-aix.fr

10.21437/Interspeech.2011-222

Abstract

We present an experiment where voice quality of French and Italian dysphonic speakers was evaluated by French and Italian listeners, specialists in phoniatrics. Results showed that both groups of speakers were perceived in the same way by the two groups of listeners in term of overall severity and breathiness. But the perception of roughness is clearly language dependant. Italian listeners underestimate roughness compare to French listeners. If we link these results obtained in perception with measures obtained in speech production, we can make the hypothesis that it is a case of perception/production adaptation process.

Keys: perception, voice quality, dysphonia, cross-language.

1. Introduction

The aim of this study is to compare the perception of voice quality among two languages: French and Italian. We studied this point in the clinical context of dysphonia. Continuous speech is considered as the most natural way (compared to sustained vowels) to evaluate the voice quality of dysphonic speakers. However, this elocution involves many linguistic, phonetic and cultural phenomena which can disturb or mask the perception of the dysfunction itself. Our objective was to test, in a crossed way, French and Italian dysphonic speakers by French and Italian listeners (specialist in phoniatrics, speech therapists) in order to evaluate if the language of the speakers and the language of the listeners can influence the result of the voice quality evaluation.

Generally, dysphonia is seen as the result of a biomechanical problem (ex: laryngeal paralysis), a physiological problem (ex: oedema) or it is considered as the result of a dysfunctional disorder (ex: vocal abuse). The main features associated to dysphonia are hoarseness, breathiness or roughness. These features are explicitly defined in the GRBAS scale of Hirano [1], where G is the degree of hoarseness (or the global severity), R is the grade of roughness, B is the grade of breathiness, A is the overall weakness of voice (asthenicity) and S is the "strained quality". Each of those dimensions can be graded perceptually from 0 to 3. Most of the time, dysphonia and its associated acoustico-perceptual features are considered as universal. It leads us two questions:
(1) Is the relationship between dysphonia and its universal features (hoarseness, roughness, breathiness) reversible? More precisely, if hoarseness, roughness or breathiness are detected in a voice, does it mean that it is a dysphonia ?
(2) Are these acoustico-perceptual features universal?

2. Source's variations and languages

Larynx is a set of anatomical pieces, which can be damaged or not used in a suitable way. But it is also the sound source of speech and voice is an essential element in the listener's analysis of the speaker's physical, psychological and social characteristics [2].

2.1. Phonetically determined variations

The larynx is an important element in the spoken communication chain. It has different phonological functions used differently among the languages [3]. For example, a particular pitch contour (rise, fall) may be used to contrast tones in a tone language, to contrast to lexical elements in stressed languages (ex: in Italian, papà = daddy, pàpa = pope), or to contrast a question with a response.

In term of voice quality, phonation types (modal, creaky, breathy and harsh) can be used linguistically [4]. For instance, in Mazatec (a language spoken in the south of Mexico), modal ("normal"), creaky or breathy vowels are phonological distinctive elements [5]. In this language, creaky /a̰/, breathy /a̤/ or modal /a/ are different phonemes in the same way than in French, the oral vowel /a/ is different from the nasal /ã/.

Others studies showed that in Mandarin, speakers often produce a tone pattern (Low dip tone 3) with creaky voice [6] probably to make the perception of this pattern more robust. Indeed, Grenié et al. [7] report that Mandarin listeners recognize Tone 3 faster when it is produced with creak than when produced without creak.

2.2. Para- socio-extra linguistic aspects

Voice quality plays a fundamental role in stylistic aspects. For instance, breathy voice can be used as an indicator of proximity or intimacy, hoarseness as a marker of emotional state [8].

Ní Chasaide & Gobl report that within a language or within a regional variety, voice quality features may signal social subgroups [4]. For instance, in English spoken in Edinburgh, creaky voice can be associated with upper classes whereas whispery, breathy or harsh voices are linked to a lower social status.

Copyright © 2011 ISCA

Yanushevskaya et al. studied vocal correlates of affect [9]. They demonstrated complex relationships between voice quality (breathy, whispered, lax-creaky, tense, modal), f_0 patterns and affective attributes perceived (sad-happy, intimate-formal, bored-interested…) by different groups of listeners (Irish-English, Russian, Spanish, Japanese). Some of the main results pointed to similarities among the language groups as well as some noticeable cross-language/culture differences in how these stimuli map to affect.

Wagner & Braun presented a study analyzing the vocal parameters (F0, F0 modulation, HNR, jitter, shimmer…) measured on large groups of Italian, German and Polish male speakers [10]. Significant differences between these groups have been found showing that these languages are characterized by different prototypic voice profiles. The Polish group exhibited the highest values with respect to HNR. As far as parameters indicating vocal instability are concerned, the Italian group is very different from the other two. With respect to the perceptive domain, Wagner & Braun concluded, with prudence, that the voices of Polish speakers will be perceived as "bright", whereas those of Italian speakers will be rated as "rough".

2.3. Cross-linguistic pathological voice assessment

Cross-languages studies on pathological voice quality are rare. Anders et al. [11] studied the effects of professional background but also the effects of culture (language) on the perception of hoarseness. They found no significant differences between classes of listeners and concluded that training, professional and cultural background did not have major influence on perceptual rating.

Yamaguchi et al. [12] experimented Japanese and American listeners using the GRBAS scale. The ratings obtained from the two groups were compared to determine if the different linguistic background affected the use of the GRBAS scale. Results showed that there were no significant differences between the Japanese and American listeners in the use of the Grade, Roughness and Breathiness scales but Asthenia and Strain scales, however, were different between the two groups of listeners.

Yiu et al. [13] studied the cross-cultural differences in the perception of voice disorders by speech pathology students from Australia and Hong Kong. Listeners were asked to rate the breathy and rough qualities of synthesized voice signals in Cantonese and English. Results showed that the English stimuli were rated less severely than the Cantonese stimuli by both groups of listeners. In addition, the male Cantonese and English breathy stimuli were rated differently by the Australian and Hong Kong listeners.

Recently, Jayakumar et al. [14] measured the effect of geographical and ethnic variation on dysphonia severity index. They measured different instrumental parameters on "normal" (G0) male and female Indian speakers. Results showed noticeable differences between Indian and European population on Maximum Phonation Time, Highest fundamental frequency and the linear combination of Dysphonia Severity Index values. Authors focused the discussion on physiological explanations and concluded by cautioning “voice professionals to reinvestigate and establish their own norms for their geographical and ethnic groups.”

2.4. What to conclude?

In speech sciences, results show that voice quality and particularly its perception is clearly language dependant. Results in clinical studies are not clear. The various utterances used, the languages involved or the various evaluation methods can explain the lack of consensus. The necessity of medical standardization is maybe a factor which favors the concept of universality for voice quality and dysphonia features.

One important point to focus is that some phonation types can be considered as a part of the communication chain in a language or as a disorder in another one. But even if we consider languages where phonation is normally modal, are we sure that some subtle effects cannot be observed on voice quality? And if a significant variation were observed between two nearby languages, it could be finally worrying in other forms of more visible variations in clinical practice such as regional variations or sociolinguistic particularities.

To obtain a part of response to these questions, we studied perception of voice quality in a clinical context across two languages: French and Italian.

3. Material and methods

3.1. Corpus

The corpus used in this experiment was the voices of native French speakers (spkFRA) and native Italian speakers (spkITA). All speakers were male (M) or female (F) adults.

The set of Italian voices (20 F, 7 M) were recorded at the Otolaryngology Dept, Ospedale Maggiore Policlinico, Milano, Italy, with the EVA2 device [15]. Patients’ pathologies were nodules (4), cysts (3), polyps (4), Reinke’s oedema (3), laryngeal paralysis (1), dysplastic lesion (2), incomplete vocal folds closure (2), synechia (1). Seven normal speakers were added. The speaking task was reading text “Il deserto”.

The set of French voices (34F, 6M) were extracted from the MTO database recorded in the ENT Department of the Timone University Hospital in Marseille, France [15]. Patients’ pathologies were nodules (8), polyps (8), Reinke’s oedema (8) laryngeal paralysis (8) and dysfonctionnal dysphonia with normal larynx (8). These voices were selected in order to correspond approximately to the Italian set in terms of gender and pathologies The second criteria in the query was to select a uniform palette from normal voice G0 to severe perturbation G3, information available in the database system management. The speaking task was reading text “La chèvre de M. Seguin”.

3.2. Listeners and perception task

The participants to the perception test were all specialists in voice therapy (ENT, phoniatricians, speech therapists). 6 Italian listeners (3 from Milano, 3 from Ravenna) and 6 French listeners (3 from Marseille, 3 from Toulouse) participated to the experiment. We wanted to have at least 2 different hospital centers per country in order to minimize some local particularities in term of professional background or practice.

The task was the perceptual evaluation GRB of Hirano [1]. For each trial, the participant listened to the voice (several times if he wanted). For each dimension G (overall severity), R (roughness) and B (breathiness), the listener gave a note: 0 if normal, 1 if slightly impaired, 2 if moderately impaired, 3 if extremely impaired. All participants were familiar with this task. The test was individual. A training phase was proposed in order to familiarize each participant to the experimental environment. Stimuli were submitted by block: first, French speakers; second Italian speakers. The order of presentation was randomized in a block. The automatic presentation of stimuli and the computerized acquisition of answers were

monitored by the software PERCEVAL with LANCELOT extension [16].

4. Results

4.1. Data processing

For the statistical analysis, we used 'R' software version 2.12 (www.r-project.org). Results are based on 6+6 listeners * 40+27 speakers * 3 dimensions GRB = 2412 perceptual tests.

It is well known that in this kind of perceptual assessment, an important variability inter listeners is observed [17]. In order to obtain a robust and reliable measure, we kept, for each group of listeners, the modal value of the group. This principle was also adopted by [12]. Modal value is the most frequent value, which is a sort of filtering by majority vote (ex: 1,1,2 => 1). It is similar to the well-known consensus method but in our experiment, the consensus was obtained by a statistical a posteriori majority vote. In case of impossibility of consensus (ex: 3, 1, 2), the data was declared non reliable.

To obtain a statistical measure of inter-group agreement for categorical items, we used Cohen's kappa coefficient with the statistical significance proposed by Landis and Koch, who characterized values < 0 as indicating no agreement and 0–.20 as slight, .21–.40 as fair, .41–.60 as moderate, .61–.80 as substantial, and .81–1 as almost perfect agreement.

4.2. Inter-group agreement

In a first step, we compared the agreement between experimental centers per country (cf. §3.2): Marseille [MRS] vs Toulouse [TLS] and Milano [MIL] vs Ravenna [RAV]. We also merge data per country. Our preliminary results, based on consensus method, showed that (Table 1)

1) Agreement was better if consensus method was applied with all the listeners of a country than if we took into account only listeners of a center

2) Agreement for G was better than agreement for R or B

Table 1: *Kappa Cohen coeff according to the groups*

	MIL vs RAV	MRS vs TLS	FRA vs ITA
G	0.48	0.66	0.78
R	0.37	0.49	0.60
B	0.40	0.49	0.55

When applied on a single experimental centre, the consensus method was based only on 3 listeners which was insufficient to obtain robust results. In particular, it was not rare to obtain ambiguous decision (ex: 3, 1, 2). It was no more the case when we merged the 6 French listeners vs the 6 Italian listeners. We finally decided to retain results per country, without distinction of the experimental centre.

4.3. Effect of speakers' language

No significant effect was found depending on the language of speakers. In other words, the 2 groups of speakers were perceived as a single group. We studied then the agreement between our two groups of listeners (listenFRA or listenITA) independently of the speakers' language.

4.4. Inter-language agreement

Concerning perception of Global severity of dysphonia, Kappa coefficient K= 0.78 indicates a substantial agreement between French and Italian listeners (Table 1). Contingency table is in Table 2.

Concerning perception of Roughness, Kappa coefficient K= 0.60 indicates a moderate agreement between French and Italian listeners (Table 1). Contingency table is in Table 3.

Concerning perception of Breathiness, Kappa coefficient K= 0.55 indicates a moderate agreement between French and Italian listeners (Table 1). Contingency table is in Table 4.

Table 2. *Contingency table in perception of G*

ITA\FRA	G0	G1	G2	G3	Sum
G0	18	2	0	0	20
G1	2	20	2	0	24
G2	0	0	14	1	15
G3	0	0	1	7	8
Sum	20	22	17	8	67

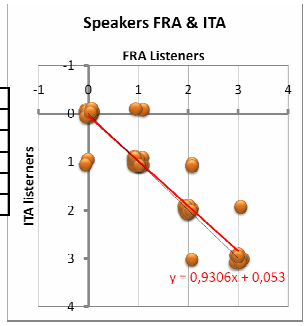


Table 3. *Contingency table in perception of R*

ITA\FRA	R0	R1	R2	R3	Sum
R0	21	6	3	0	30
R1	4	15	3	2	24
R2	0	0	11	0	11
R3	0	0	0	2	2
Sum	25	21	17	4	67

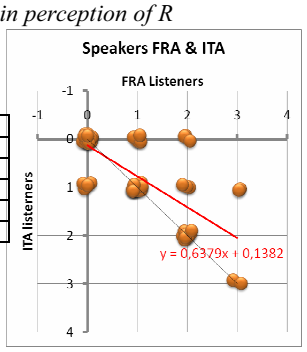
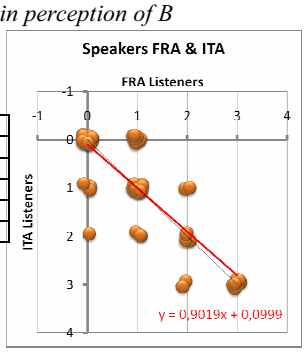


Table 4. *Contingency table in perception of B*

ITA\FRA	B0	B1	B2	B3	Sum
B0	24	8	0	0	32
B1	4	13	2	0	19
B2	1	3	4	0	8
B3	0	0	2	6	8
Sum	29	24	8	6	67



4.5. Randomized or organized variation?

In order to evaluate if inter-group moderate agreement is only a question of variability or if we can find tendencies between languages of listeners, we used a linear correlation model

`lm(ITA~FRA, data = G or R or B)`

Correlation between French and Italian listeners is pretty correct for G (R^2 = 0.88). The regression coefficient is a=0.93 (Table 2). Intercept is not significant.

Correlation between French and Italian listeners is moderate for R (R^2= 0.52). The regression coefficient is a=0.64 (Table 3). Intercept is not significant.

Correlation between French and Italian listeners is moderate for B (R^2= 0.68). The regression coefficient is a=0.90 (Table 4). Intercept is not significant.

5. Discussion

We did not found significant effect of the speakers' language, neither interaction between speaker's or listener's languages.

We can explain this result by the fact that the perception task was in the frame of a clinical context, where the attention is focused on non linguistic features. Moreover, a part of the French listeners were familiar with Italian and a part of the Italian listeners were familiar with French. It is possible that during this perception task, these listeners used the same cognitive mechanisms quite independently of the speaker's language. Results are more interesting in term of distinction between the two groups of listeners.

The perception of the dysphonia overall severity is not language dependant. The consistency is good and the agreement substantial. In fact, this dimension is essentially a question of quantity. Italian and French specialists have globally the same references.

The perception of breathiness is quite similar even if agreement is lower. The regression coefficient indicates that Italian listeners are slightly more indulgent than French ones.

The main effect that we have observed concerned the perception of roughness. Italian listeners are clearly more indulgent than French. Over 67 voices, 14 stimuli were underestimated by Italian specialists compared to French ones. 5 of these stimuli were quoted with a difference of 2 degrees. If we remove R0 data which cannot be underestimated, we obtain 33% of data which were underestimated. Of course, we could also interpret that French listeners are more severe than Italians and overestimated the roughness compared to Italians. It depends on the reference that we take to make the comparison. However, these results can be linked to the study of Wagner & Braun [10] mentioned above in §2.2 where they observed that parameters indicating vocal instability in the Italian group is significantly more important compared to other languages. For us, it is a pity that this study did not include French in the cross-language comparison because we could have a direct comparison between French and Italian in term of acoustical parameters. But the authors finally conclude, with prudence, that the voices of Italian speakers will be rated as "rough". Recall that the population tested in the study of Wagner & Braun was normomorphic speakers. Consequently, we can make the hypothesis that in Italy, a noticeable level of roughness (compared to other languages) is present even in normomorphic speakers. Therefore, to be perceived as abnormally and clinically rough, the level of roughness should be higher. Consequently, if voice therapists have naturally adapted their thresholds to the local situation, they are naturally more tolerant to roughness, result that we have observed in our experiment. We can wonder why Italian speakers have eventually a prototypic vocal instability higher than in other languages (and it will be interesting to examine precisely this point)? Because listeners are more tolerant to roughness. We can make the hypothesis that this is a typical case of perception/ production adaptation process.

6. Conclusion

To the question "Is the relationship reversible between dysphonia and its universal features?", we can conclude negatively. Hoarseness, roughness or breathiness are parts of the communication chain as normal phenomena. The issue is finally to find the limits between normal functioning and an excessive process linked to a disorder. To the question "Are these acoustico-perceptual features universal ?" Probably not. Voice quality and especially its perception is language dependent. Even if the necessity of standardization in clinical context is a legitimate goal, adaptation should be considered as proposed by [14]. Problem of cross linguistic pathological voice database can be also an important issue as mentioned in [15]. Maybe these cross-linguistic aspects can be seen as minor in a clinical point of view, except in multilingual countries, but it can reveal more serious issues if we consider other forms of more common variations in clinical practice such as regional variations or sociolinguistic particularities.

7. Acknowledgements

We thank MD.Guarella, C.Spezza, M.Puech, S.Cretani., T. Fuschini., V.Geminiani., A.Zambarbieri for their participation. This research was partially supported by COST Action 2103 “Advanced Voice Function Assessment” and ANR BLAN08-0125 of the French National Research Agency.

8. References

[1] M. Hirano, Clinical Examination of Voice. Wien: Springer Verlag, 1981.

[2] J. Laver, The phonetic description of voice quality. Cambridge University Press, 1980.

[3] J. Vaissière, "Phonological Use of the Larynx", Proc. Larynx 1997, Marseille, France, 1997, 115-126.

[4] A. Ní Chasaide, A. Gobl, "Voice Source Variation", in The Handbook of Phonetic Sciences, W. J. Hardcastle, J. Laver, Ed. Oxford, UK: Blackwell Pub. Ltd, 1999.

[5] P. Ladefoged, I. Maddieson, The sounds of the world’s languages. Blackwell Publishers, 1996.

[6] J. Kreiman, B. R. Gerratt, S.D. Khan,"Effects of native language on perception of voice quality", Journal of Phonetics,38: 4, 588-593, 2010.

[7] A. Belotel-Grenié, M. Grenié, "Types de phonation et tons en chinois standard", Cahiers de linguistique Asie orientale, 26: 2, 249-279, 1997.

[8] C. D’Alessandro, "Voice Source Parameters and Prosodic Analysis", in Methods in empirical prosody research, S. Sudhoff, Ed. Walter de Gruyter, 2006

[9] I. Yanushevskaya, A. Ní Chasaide, C. Gobl, "Cross-Language Study of Vocal Correlates of Affective States", Proc. Interspeech, Brisbane, Australia, 2008, 330-333.

[10] A. Wagner, A. Braun, "is voice quality language dependant ?", Proc. ICPhS, Barcelona, 2003, 651-654

[11] L. C. Anders, H. Hollien, P. Hurme, A. Sonninen, J. Wendler, "Perception of Hoarseness by Several Classes of Listeners", Folia Phoniatr Logop, 40:2, 91-100, 1988

[12] H. Yamaguchi, R. Shrivastav, M. L. Andrews, S. Niimi, "A Comparison of Voice Quality Ratings Made by Japanese and American Listeners Using the GRBAS Scale", Folia Phoniatr Logop, 55: 3, 147-157, 2003.

[13] E. M.-L. Yiu, B. Murdoch, K. Hird, P. Lau, E. M. Ho, "Cultural and language differences in voice quality perception: a preliminary investigation using synthesized signals", Folia Phoniatr Logop, 60: 3, 107-119, 2008.

[14] T. Jayakumar, S. R. Savithri, "Effect of Geographical and Ethnic Variation on Dysphonia Severity Index: A Study of Indian Population", J Voice, in Press, 2010

[15] A. Ghio, G. Pouchoulin, B. Teston et al., "How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers?", Speech Comm, Accepted In Special Issue "Advanced Voice Assessment", 2011.

[16] C. André, A. Ghio, C. Cavé, B. Teston, "PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception", Proc. ICPhS, Barcelona, Spain, 2003, 1421-1424.

[17] J. Kreiman, B. R. Gerratt, G. B. Kempster, A. Erman, et G. S. Berke, "Perceptual evaluation of voice quality: review, tutorial, and a framework for future research", J Speech Hear Res, 36:1, 21-40, 1993.