Top-Down Mechanisms in Dysphonia Perception: The Need for Blind Tests

Alain Ghio, Joana Révis, Sabine Merienne, Antoine Giovanni

▶ **To cite this version:**

Alain Ghio, Joana Révis, Sabine Merienne, Antoine Giovanni. Top-Down Mechanisms in Dysphonia Perception: The Need for Blind Tests. Journal of Voice, 2013, 27 (4), pp.481-485. 10.1016/j.jvoice.2013.03.015 . hal-01486663

HAL Id: hal-01486663
https://hal.science/hal-01486663v1

Submitted on 20 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Top-Down Mechanisms in Dysphonia Perception: The Need for Blind Tests

***Alain Ghio, *,†Joana Révis, †Sabine Merienne, and *,†Antoine Giovanni,** **Aix-en-Provence, and †Marseille, France*

Summary: The purpose of this study was to determine the extent to which the information a therapist or a physician has about a dysphonic speaker, particularly whether he or she is in the pretreatment or posttreatment period, can influence judgments of the patient's voice. The voices of 53 dysphonic speakers were used in the study. For each speaker, we selected a pair of voice samples recorded under different circumstances. Seven listeners who were speech therapists, ear, nose, and throat surgeons, or voice pathologists took blind-listening tests in which they were asked to compare the two voices in each pair (phase 1: blind listening). A few weeks later, the listeners took the very same test again, except that this time, they were given bogus information about whether the speaker had/had not been treated by laryngeal surgery or speech therapy (phase 2: influenced listening). The information given for each voice sample either reinforced the judgment made in phase 1 (eg, the voice judged to be better on the blind test was said to be posttreatment) or countered that judgment (eg, the voice rated as better on the blind test was said to be pretreatment). The influenced-listening results showed that in the reinforced condition, the original ratings were significantly amplified. By contrast, in the countering-influence condition, decision changes were frequent: we found that judgment reversals and the countering-information scores were almost independent of the blind-listening scores. These findings point out the dire need to use a blind protocol in perceptual assessments of dysphonia.

Key Words: Dysphonia–Voice assessment–Perception–Top-down processing–Bottom-up processing.

INTRODUCTION

Perception of voice quality

In treating dysphonia, a perceptual assessment of the quality of the patient's voice is conducted to provide clinical information about the type and severity of the dysfunction. This is generally done using a standard scale containing several parameters that listeners must judge by ear. Hirano's[1] Grade, Roughness, Breathiness, Asthenia, Strain (GRBAS) is the most commonly used scale. In clinical practice, perceptual assessment is considered the gold standard for rating voice quality. However, although the GRBAS is the most widespread scale in use today, and although attentive listening makes an undeniable contribution to painting a complete clinical picture of the patient, the actual utility of this scale as a reliable assessment device is a subject of regular debate. Many studies have found clear evidence of variability in judgments of the same voice made by different listeners and in judgments made by the same listener at different times.[2–7]

This lack of reliability can be explained in terms of the context sensitivity of speech-perception mechanisms. Gerratt et al[8] showed that the context in which voice samples are presented can affect judgments. Martens et al[9] found that reading the voice spectrogram while listening to a voice increased between-listener reliability of voice-quality judgments. These findings illustrate that perception-based decisions are complex. In the study by Martens et al, the judgments not only called on the auditory system per se but were also based on visual information drawn from the spectrogram and the knowledge the listeners possessed about how to interpret the spectrotemporal representation. Judgment variability then—a reflection of poor instrument reliability—has already been observed under controlled experimental conditions. But what happens in daily clinical practices? How can one account for this variability?

Importance of top-down processing in perception

Evaluating a voice by ear consists of interpreting the sound signal heard at a given moment. The risk incurred in this kind of assessment is that the outcome may differ across listeners, depending on their individual listening habits, and over time, depending on the information currently available to the listener. Rating the voice quality of a speaker is mainly a bottom-up perception process. That is, based on the voice sample heard, listeners categorize the voice by interpreting acoustic cues detected perceptually. But like all other speech perception processes, it cannot be reduced to this simple bottom-up route that goes from the acoustic to the cognitive. Top-down processing also takes place and influences the listener's perception. For example, when we hear a degraded, noisy, or phonetically impoverished utterance, top-down processes attempt to restore the degraded parts of the signal and optimize message intelligibility. Attention allocated to the message will maximize or minimize the effects of the restoration process.[10]

In the domain of visual perception, Simons and Chabris[11] showed that when our attention is focused on another task, or on another object present in a visual scene, we can be blind to certain salient unexpected elements of the scene. They called this phenomenon "inattentional blindness." In the experiment by Vitevitch,[12] participants were instructed to repeat words that varied in lexical complexity. In the middle of the list, the voice used to produce the to-be-repeated words sometimes changed. At least 40% of the participants did not detect the speaker change.

Our sense of smell is also subject to distortion effects and perceptual illusions. Language, for example, can interfere

From the *Aix-Marseille University, CNRS, Laboratoire Parole et Langage, Aix-en-Provence, France; and the †Department of Otolaryngology-Head and Neck Surgery, La Timone Universitary Hospital Center, Marseille, France.

Address correspondence and reprint requests to Alain Ghio, Aix-Marseille University, CNRS, Laboratoire Parole et Langage, 5 Avenue Pasteur, BP 80975, 13604 Aix-en-Provence Cedex 1, France. E-mail: alain.ghio@lpl-aix.fr

Journal of Voice, Vol. 27, No. 4, pp. 481-485
0892-1997/$36.00
© 2013 The Voice Foundation
http://dx.doi.org/10.1016/j.jvoice.2013.03.015

with this perceptual modality, as noted by Herz[13] in a study where the verbal context influenced the perception of smells. The mere fact of associating a verbal label to an odor was likely to generate an olfactory illusion: one and the same odor could be judged differently, depending on what name it was given.

Hypotheses

We hypothesized that the phenomena described previously play a part in decision making during judgments of dysphonia. Top-down mechanisms may mask salient acoustic events or restore phenomena that do not exist. And the listener's selective attention may make him or her deaf to certain acoustic realities. For example, in the case of laryngeal paralysis, listeners may focus on vocal aspects related to glottal leakage (breathy voice) and may not hear other normal or dysfunctional features of the voice. Similarly, listeners may be insensitive to change. Last, verbal information given during listening may substantially modify judgments of voice quality. More specifically in the present study, we hypothesized that a listener's knowledge of whether a patient is in the pretreatment or posttreatment period would modify his or her perceptual ratings of the quality of the patient's voice.

MATERIALS AND METHOD

Experimental design

The experimental procedure consisted of having a panel of experienced judges who blindly rate dysphonic voices (phase 1) and then rate them again later when given information about the treatment allegedly undergone by the patient (phase 2). If the experiment is methodologically sound, then any differences between the ratings given on the two phases can be attributed to the information given to listeners.

Voice samples were presented in pairs, each pair being made up of two samples from the same speaker (hereafter called voice A and voice B). Given that the two recordings were made on different dates, the voice quality of the two samples was usually different (pretherapy/posttherapy, recordings spaced out over time …). The listeners had to compare the two voices in each pair after listening to the pair several times if desired. Dysphonia rated was made on a seven-point comparative scale: voice A is much less dysphonic, less dysphonic, slightly less dysphonic, equally dysphonic, slightly more dysphonic, more dysphonic, much more dysphonic than or as voice B.

We chose this scale to put the listener in conditions like those found in clinical practice, where the main purpose of the assessment is often to perceive the amount of change brought about by a given treatment (improvement or deterioration) and to obtain a sufficiently sensitive measure. We could have used a traditional scale like the GRBAS,[1] but a four-level absolute rating scale seemed too insensitive to slight differences in quality, as compared with our seven-level comparative test.

Listeners

All participants were accustomed to hearing dysphonic voices. They included three ear, nose, and throat (ENT) surgeons, three speech therapists, and one voice pathologist. Of course, to avoid biasing the experimental results, the participants were unaware of the true purpose of the study, which was claimed to be aimed at testing a computerized protocol for rating dysphonia in hospitals.

Corpus

The recordings proposed to the listeners were selected from the dysphonic speaker database compiled at the ENT department of the Timone University Hospital in Marseille, France.[14] The 53 patients selected as speakers for the study were adults with vocal fold nodules or polyps (44 women and 9 men). We limited the study to these two pathologies, not only to reduce the amount of heterogeneity in the forms of expression of dysphonia but also because they could be treated by surgery and/or speech therapy, a necessary condition for the second phase of the experiment. To obtain voice pairs for each speaker who would fit with our experimental design, we selected patients for whom we had at least two recordings made on different dates.

The speakers read several passages taken from the first chapter of "La chèvre de Monsieur Seguin" (Mr Seguin's Goat), a tale by Alphonse Daudet. The mean duration of the utterances was 20 seconds.

Procedure

Experimental setup and methodological precautions. The experiment was run using PERCEVAL freeware with its LANCELOT extension (www.lpl-aix.fr/∼lpldev/perceval). The listening sessions were conducted in a closed room, on the same computer and with the same soundboard and the same headphones in each session. The experiment included four sessions: two blind-listening sessions (test-retest) followed by two influenced-listening sessions (test-retest). In the blind condition, the listeners had no information about the voices of the speakers they were listening to. In the influenced-listening sessions, bogus information was displayed on the screen stating the patient's type of treatment (surgery or speech therapy), and for each voice sample, whether it was recorded pretreatment or posttreatment. For each session, the instructions were displayed in writing on the computer screen. Before beginning the actual test, three practice items were proposed to familiarize each listener with the task and rating scale. Each session consisted of two blocks of 16–25 voice pairs, with a 5-minute break between blocks to reduce fatigue and inattentiveness. Half of the listeners began with one block, and the other half began with the other block. In each block, the pairs were presented in random order so as to minimize the list effect. Last, for each listener, the listening sessions were held at least 1 week apart to prevent memorization.

Test-retest. Each listening condition (blind or influenced) consisted of a test and a retest, with stimulus order varied across listeners, and between the test and retest sessions for the same listener. Having participants perform the tests twice (test, retest) allowed us to compute averages for both types of listening and thus decrease some of the effects of random errors (listener distraction, poor manipulation of response choices, etc.), as suggested by Bele.[6]

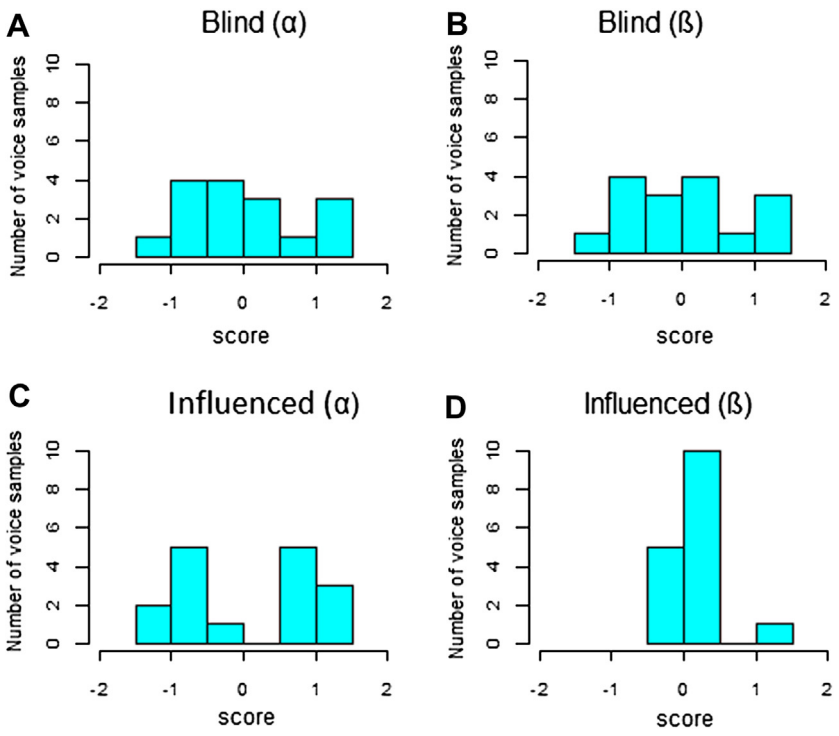

FIGURE 1. Results of the blind-listening (**A**, **B**) and influenced-listening (**C**, **D**) tests. In influenced listening, the α and β data sets were used to supply consistent or inconsistent contextual information, respectively.

Setting up the influenced-listening sessions. The blind-listening test gave us 14 ratings for each voice pair (seven listeners × two sessions). Each rating was converted into a score as follows: 3 (much less dysphonic), 2 (less dysphonic), 1 (slightly less dysphonic), 0 (equally dysphonic), −1 (slightly more dysphonic), −2 (more dysphonic), and −3 (much more dysphonic). The mean of the 14 scores was used to rank the pairs in decreasing order of voice quality. The closer the absolute value of the mean was to three, the greater the amount of change between voice A and voice B (with +3 reflecting a very good quality for A and −3 a clear preference for B). The closer the absolute value of the mean was to zero, the more the listeners felt that voices A and B were equivalent in terms of voice quality.

The data obtained from this first phase of the experiment were divided into two parts with similar distributions, hereafter called data sets α and β. This was done by applying the following simple principle: data set α was made up of pairs in positions 1, 3, 5, 7, … 45, 47, 49 of the ranking; data set β was made up of pairs in positions 2, 4, 6, 8, … 46, 48, 50.

For data set α, the information was consistent (in the clinical sense of the term) insofar as those voices rated as less dysphonic on the blind test were said to be after treatment. For data set β, the information given on the second phase was inconsistent insofar as the voices judged to be more dysphonic on the blind test were said to be after treatment. Note that the information that accompanied the voices on phase 2 of the experiment did not reflect the patients' actual pretreatment or posttreatment situation but was fabricated so as to obtain a perfect balance across testing conditions and a symmetrical experimental design.

To avoid having the inconsistent information in data set β seem too unrealistic, which could raise some doubts in the minds of listeners, we eliminated those voice pairs whose mean score on the blind test was extreme, that is, too large a difference in quality between voice A and voice B. In these cases, the voices rated as much more dysphonic would have been said to be posttreatment in data set β, which voice therapists would find highly unlikely and thus difficult to accept. For both data sets (α and β), we selected pairs whose mean score on the blind test fell between +1.5 and −1.5, which gave us 32 pairs. Last, to prevent a listening-order effect, the voice sample in each pair said to be pretreatment was sometimes heard first and sometimes heard second. This was aimed at reducing the recency effect (better memory trace for the last voice heard).

RESULTS

Statistical analyses were run under "R" software version 2.12.0 (www.r-project.org). Whether in blind or influenced listening, the score retained for each voice pair was the score averaged over 14 judgments (seven listeners × two listenings for each condition). In all, the experiment included 700 voice-pair listenings (14 judgments × 50 pairs) for the blind rating and 448 (14 × 32) for the influenced rating, that is, 2296 listenings.

Blind listening

The scores obtained on the 50 voice pairs in blind listening were between −3 and +3. As stated previously, extreme scores were discarded to avoid having to supply overly improbable countering information on the second phase. The rest of the corpus

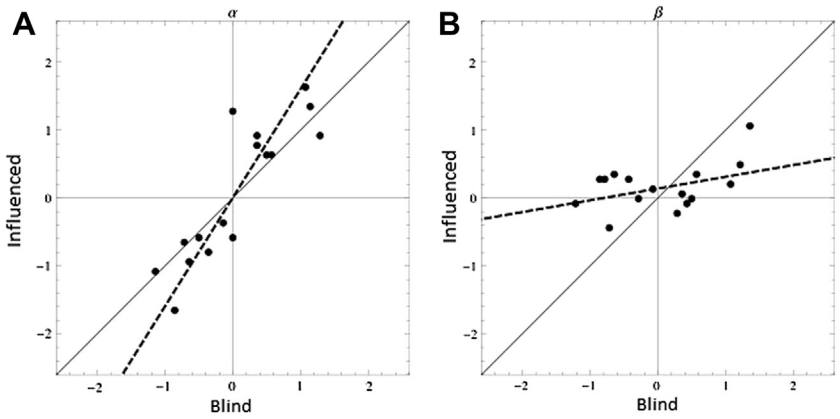

FIGURE 2. Results of influenced listening (vertical) as a function of blind listening (horizontal). (**A**) Provision of consistent information (data set α). (**B**) Provision of inconsistent information (data set β). The solid line corresponds to the absence of an effect (same scores on the blind and influenced tests). The dotted line is the linear regression line obtained from the observed scores.

(32 voice pairs) was split into two parts with similar distributions. No significant difference was found between the two data sets ($F(1,30) = 0.0011$, $P = 0.97$).

Influenced listening

Again, in the influenced-listening condition, listeners were given written information that was either clinically consistent (α) insofar as the voices judged less dysphonic during blind listening were said to be posttreatment or clinically inconsistent (β) insofar as the voices judged more dysphonic during blind listening were said to be posttreatment. Obviously, during this listening test, the stimuli in data sets α or β were mixed together and presented randomly. The score distribution for the 32 voice pairs is shown in Figure 1.

In Figure 1, we can see a pronounced effect of contextual information during influenced listening. In the condition where the information was consistent, the distribution was nearly bimodal (Figure 1, C), which corresponds to clear-cut decisions. In the condition where the information was inconsistent, the ratings were close do zero (equal voice quality for the two voices in a pair), which corresponds to nondifferentiation (Figure 1, D).

We can hypothesize that the provision of consistent information had an amplifying effect: voices rated as slightly less dysphonic on the blind-listening test were judged clearly less dysphonic because of the fact that the patients were said to be in the posttreatment phase; voices rated as slightly more dysphonic on the blind-listening test were judged clearly more dysphonic because of the fact that the patients were said to be in the pretreatment phase. This hypothesis can account for the bimodality of the distribution in Figure 1.

By contrast, in the inconsistent condition, we can hypothesize that providing countering information had a reducing effect: voices rated as slightly less dysphonic on the blind-listening test were judged to be of equal quality as the other voice in the pair as they were said to be in pretreatment, and voices rated as slightly more dysphonic on the blind-listening test were judged to be of equal quality as the other voice in the pair as they were said to be in posttreatment.

To verify these hypotheses, we conducted a linear regression analysis of the scores obtained in influenced listening as a function of the scores obtained in blind listening, for each data set (α and β). The results are shown in Figure 2.

The regression line (y-intercept at the origin, slope of +1) indicates the absence of a contextual-information effect. Points on this line correspond to ratings where scores in influenced and blind listening were equal. The distribution of scores with respect to this line thus supplies important information.

In condition α, where consistent information was given, the statistical analysis indicated a regression line slope of 1.60 ± 0.19. This slope was obtained using the weighted least-squares method while simultaneously taking into account the uncertainty levels of the blind and influenced data. Each score (blind or influenced) was assigned a weight that was inversely proportional to the interrater variability rate (standard error). The steep slope of 1.6 validates the hypothesized amplifying effect of the consistent context: the score increased by 60% over that obtained in blind listening.

In condition β, where inconsistent information was given, the regression line slope was 0.17 ± 0.14, which is low, and thus validates the hypothesized reducing effect of this condition. Recall that a null slope would indicate that the ratings made with contextual information were totally independent of the ones given in blind listening. The low value of 0.17 means that listeners given information that contradicted their perception tended to base their ratings on the contextual information, which strongly reduced the effects perceived on the blind test. In fact, we can even see some reversed judgments (points located in the upper-left or lower-right quadrant of Figure 2, B). These reversals represent 50% of the cases in the inconsistent situation; 100% of these reversals were affected by the contextual information given to participants.

DISCUSSION

In the consistent condition, the information provided had an amplifying effect. This result is in line with the findings obtained by Herz[13] for the perception of odors, where judgments were usually amplified by the association of verbal information

to an olfactory stimulus. In this condition of our experiment, the judgment context was provided by giving listeners consistent information: voice quality was better after treatment than before.

Our findings for the inconsistent condition were also analogous to those obtained by Herz, who noted that for certain odors, under the sole effect of the verbal context, up to 88% of the subjects made a perceptual interpretation that was totally different in the two sessions (positive vs negative connotation). In our experiment, we obtained judgment reversals in 50% of the inconsistent cases.

These results confirm the idea that perception is a mental construction based on the processing of available information. In this study, the auditory stimuli were exactly the same in the two listening conditions. The only thing that differed was the information given to the listeners about the speaker. This information resulted in substantial differences in the ratings. We can conclude, then, that therapists' perceptions are influenced by top-down cognitive processing (a posttreatment voice is better than a pretreatment one) and that such information makes them deaf to phenomena perceived during blind testing. We can also interpret this phenomenon as an attention attractor (I'm told that the patient has been treated, and I hear only what I expect to hear: an improvement).

Our listeners were voice-care professionals. As such, they were, and legitimately so, in a situation with strong implications in terms of therapeutic success. It would be worthwhile to carry out similar experiments on listeners who are not concerned with this issue, to see whether the effects would still occur or would disappear. In addition, we chose ENT surgeons as well as speech therapists and pathologists as listeners in our study, in the light of the fact that during the influenced-listening test, we manipulated not only the pretreatment/posttreatment situation but also the type of treatment (surgery or speech therapy). We wanted to find out if the contextual-information phenomenon varied according to the listener's profession. For example, we hypothesized that the speech therapists would be more sensitive to treatment via speech therapy than via surgery. The small number of listeners per group (three in each group) did not allow us to measure any potential group effects. This could be done in the future with a larger number of listeners.

CONCLUSION

Evaluating the outcome of treatment for dysphonia is a critical task for voice pathologists and therapists. Does surgery or speech therapy have a positive, negative, or negligible effect? Careful listening to the voice before and after treatment is one way of answering this question. But is this approach a valid one, given that the therapist or surgeon who is assessing the voice has a great deal of information about the patient's medical past and is sometimes even judging his/her own work? All the findings of our study stress the dire need to use only blind assessments for perceptual ratings of dysphonia.

Acknowledgments

We thank the French National Research Agency (ANR) for funds allocated to this study as part of the DESPHO-APADY ANR-08-BLAN-0125 research project, which enabled the compilation and exploitation of the pathological speech corpus used in this study.

REFERENCES

1. Hirano M. *Clinical Examination of Voice*. Vienna, Austria: Springer-Verlag; 1981.
2. Kreiman J, Gerrat BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Hear Res*. 1992;35:512–520.
3. Kreiman J, Gerrat BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res*. 1993;36:21–40.
4. De Bodt MS, Wuyts FL, Van De Heyning PH, Croux C. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997;11:74–80.
5. Révis J, Giovanni A, Wuyts F, Triglia J. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop*. 1999;51:108–116.
6. Bele I. Reliability in perceptual analysis of voice quality. *J Voice*. 2005;19:555–573.
7. Shrivastav R. Multidimensional scaling of breathy voice quality: individual differences in perception. *J Voice*. 2006;20:211–222.
8. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. *J Speech Hear Res*. 1993;36:14–20.
9. Martens JW, Versnel H, Dejonckere PH. The effect of visible speech in the perceptual rating of pathological voices. *Arch Otolaryngol Head Neck Surg*. 2007;133:178–185.
10. Warren RM, Warren RP. Auditory illusions and confusions. *Sci Am*. 1970;223:30–36.
11. Simons DJ, Chabris CF. Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*. 1999;28:1059–1074.
12. Vitevitch MS. Change deafness: the inability to detect changes between two voices. *J Exp Psychol Hum Percept Perform*. 2003;29:333–342.
13. Herz RS. The effect of verbal context on olfactory perception. *J Exp Psychol Gen*. 2003;132:595–606.
14. Ghio A, Pouchoulin G, Teston B, et al. How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers? *Speech Commun*. 2012;54:664–679.