

Comparison of Two Perceptual Methods for the Evaluation of Vowel Perturbation Produced by Jitter

*Jorge A. Gurlekian, *Humberto M. Torres, and †María E. Vaccari, *†Buenos Aires, Argentina

Summary: Objectives. To explore perceptual evaluation of jitter produced by fundamental frequency ($F0$) variation in a sustained vowel /a/, using two different methods. One is based on listener's internal references and the other is based on external references provided by the experimenter.

Methods. We used two methods: one is magnitude estimation-converging limits (ME-CL), which is close to the standard approach used by speech therapists when they use numerical estimations and their own standards, and other is intramodal matching procedure (IMP), where each matched stimulus is to be compared with a fixed-set matching stimuli. Systematic variations were introduced in vowel /a/ by Linear Prediction Coding synthesis using an $F0$ contour function obtained from a statistical jitter model. Six jitter values were used for each of two reference $F0$ values. Three groups of listeners were tested: expert speech therapists, speech therapy students, and naïve listeners.

Results. Perceptual functions appear to be similar and linear for both methods as the theory predicts. The answers of all groups of listeners tested with ME-CL present higher standard deviations than for IMP. When subjects were tested with IMP, intrareliability and interreliability measurements show a significant improvement for both expert and naïve listeners.

Conclusions. Both intraindividual and interindividual differences for expert speech therapists could be better managed when tested with an IMP than when they use numerical estimations and internal standards to evaluate vowel perturbation produced by jitter. This procedure could be the basis for the development of a clinical evaluation tool.

Key Words: Perception–Voice evaluation–Magnitude estimation–Matching.

INTRODUCTION

The lack of agreement and reliability regarding the perceptual evaluation of dysphonic voices is well known among speech therapists, in particular when variations of fundamental frequency (jitter) and amplitude (shimmer) occur.¹ Experts at the clinic make use of their internal references to judge perturbation. One approach for training young speech pathologists is that they develop their own internal standard earned with the experience of years of listening. Experts noted that when this training is in progress, internal references may also change over time influenced by the level of severity of their patients. To perform this task in a reliable manner, the professional must develop a kind of absolute ear for frequency perturbation over time. But, even if they accomplish the task with repeatability, reaching interrater consensus is much less probable because each professional develops his/her own reference. Perceptual judgments at the clinic, using a variety of experimental tasks such as visual analog² and magnitude estimation such as RBH (Rauheit/Roughness, Behauchtheit/Breathiness, and Heisekeitsome/Hoarseness)³ or GRBAS (Grade, Roughness, Breathiness, Asthenia, and Tension),⁴ always require that listeners rely on an internal reference or scale. Regarding these methods, some doubts arise about the scale that different

clinicians are using to judge perturbation: is it a category scale or a proportion scale? Guirao⁵ introduced the method of Magnitude Estimation-Converging Limits (ME-CL) indicating that this method is a compromise between both scales. Eisler and Guirao⁶ claim that this procedure allows more consistent judgments and produces less individual scatter in the data than the conventional magnitude estimation method. In this article, ME-CL method was used as one of the methods to represent the use of internal references.

Nevertheless, sensation thresholds to $F0$ variations could change according to listener's levels and types of experience. In a recent work,⁷ the effect of types of experience was shown to be more important than levels on the judgment of voice quality. The use of numerical scales and a variable remembering of the standards are also some of the conditions that contribute to the variability at the clinic, that is, new professionals with less exposure to pathologic voices are likely to have different standards. Even at the laboratory, some experienced listeners using a four- or seven-point categorical scale show that the internal standard is inherently unstable and judgments shift relative to the set of stimuli used in the experiment. The use of a reference standard could help but is not a guarantee to reduce variability. Paired comparisons did not increase reliability between listeners either.⁸ In a tutorial review article,⁹ an external reference was recommended as a possible solution to cope with the sources of variability. Following this advice, Gerratt and Kreiman¹⁰ proposed a multiparameter matching task to quantify voice quality.

For clinical practice, sorting stimuli has been proposed to avoid the use of internal references.¹¹ An interesting scheme for the purpose of clinical practice was investigated and compared with the GRBAS scale.¹² In this approach, raters

Accepted for publication May 15, 2015.

From the *Laboratorio de Investigaciones Sensoriales, INIGEM, CONICET-UBA, Buenos Aires, Argentina; and the †Facultad de Medicina, University of Buenos Aires, Fonoaudiología, Buenos Aires, Argentina.

Address correspondence and reprint requests to Jorge A. Gurlekian, Laboratorio de Investigaciones Sensoriales, INIGEM, CONICET-UBA, Av Cordoba 2351, 9 piso Sala 2, 1120 C.A.B.A., Argentina. E-mail: jag@fmed.uba.ar

Journal of Voice, Vol. ■, No. ■, pp. 1-8

0892-1997/\$36.00

© 2015 The Voice Foundation

<http://dx.doi.org/10.1016/j.jvoice.2015.05.009>

must reorder a sequence of voices with different degrees of perturbation as a way of training. Despite encouraging results as evaluators ordered pathological voices, they complained about the difficulty of order in one dimension, stimuli with complex variations of roughness and breathiness.

Psychophysical methods were compared to evaluate breathiness^{13,14} showing less dispersion for a matching-production task.

The present work explores an experimental procedure following the idea of using external standards. Our main interest is to focus in a method to improve the interrater agreement and reliability. Experiments were conducted with several external references on the basis of the matching task¹⁵ to evaluate the capacity to identify percentages of jitter using this method. Agreement between expert listener responses for matching will be compared with results obtained from the ME-CL method.

We chose jitter as one of the acoustic correlates of roughness, considering that psychoacoustic functions of roughness use frequency modulation to obtain the physical continuum.^{16,17} Another reason for this selection is that abnormal jitter values are more frequently found in normal voices when compared with shimmer¹⁸ although the presence of shimmer produces a similar effect than jitter to produce roughness sensation,⁷ and finally because only few works produced systematic jitter variations.

Experimental procedure

Stimuli. On the basis of the acoustic parameters of both female and male^{19,20} vowels /a/ indicated on Table 1, two synthesized vowel versions were created by an Linear Prediction Coding (LPC) formant synthesizer at a sampling rate of 50 KHz and 16 bits. Once the average $F0$ was selected, either 120 Hz or 240 Hz, it remained constant for all the vowel duration of 3 seconds for a jitter value of 0%. This supranormal stimulus was located at one extreme of the continuum. At the other extreme, we produced a stimulus with a jitter value of 3% according to Equation 1. Five intermediate jitter values completed the linear sequence of values with variation percentages indicated on Table 2.

Fundamental frequency variation over time was created to produce stimuli with these intended jitter values. One possible way to do this is to introduce random noise in the glottal source to create controlled variations.^{21,22} The method used in this article uses a statistical model of jitter.²³ We chose the definition of percent jitter (Equation 1) as the average of the difference

TABLE 2.
Jitter in % for Intended Values and Their Real Production and the Difference Between Them

No.	120 Hz			240 Hz		
	Planned	Real	Dif	Planned	Real	Dif
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.50	0.52	0.02	0.50	0.52	0.02
3	1.00	1.02	0.02	1.00	1.06	0.06
4	1.50	1.56	0.06	1.50	1.51	0.01
5	2.00	2.08	0.08	2.00	1.99	0.01
6	2.50	2.55	0.05	2.50	2.53	0.03
7	3.00	3.06	0.06	3.00	2.97	0.03

between two $F0$ values, normalized to the average $F0$ and multiplied by 100.

$$J\% = 100 \frac{\frac{1}{N-1} \sum_{i=1}^N |F0_i - F0_{i+1}|}{\frac{1}{N} \sum_{i=1}^N F0_i} \quad (1)$$

where $F0_i$ is the i th fundamental frequency cycle to cycle and N is the number of cycles.

By using this method for the creation of vowel stimuli, $F0$ variations are made independent from amplitude variation which remained constant. According to Titze²⁴ and Torres et al.,²⁵ $F0_i$ values have a Gaussian behavior with normal density probability functions ($F0, \sigma_{F0}$). As a result, it is possible to synthesize vowels from reference average $F0$ values and each intended jitter value. The $F0_i$ values set have a Gaussian noise distribution with an average $F0$, either 120 Hz or 240 Hz, and standard deviation given by the Equation (2).

$$\sigma_{F0} = \frac{\sqrt{\pi}}{200} F0 J\% \quad (2)$$

Once the $F0$ set which represents the glottal source is defined, stimuli are synthesized using the LPC method.²⁶ Stimuli are further verified to insure that they effectively have the programmed perturbation. Differences between programmed and real values do not exceed a maximum value of 4% as summarized in Table 2. Figure 1 presents an example of original and modified stimulus signal, with a jitter of 0 and a jitter of 3, respectively.

Test protocols

Magnitude estimation-converging limits method. For test 1, we used the ME-CL method.⁶ Six stimuli indicated in Table 2 (from 2 to 7) were presented in a random order for each listener, with four repetitions each and 4 seconds of inter-stimulus interval for a total of 24 trials. The test was implemented in a graphical software interface for orientation, testing, and response collection. Instruction to subjects was “you will hear the vowel /a/, your task is to evaluate the degree of perturbation by giving a number to each stimuli according to a numerical scale of your preference. Always assign high numbers to stimuli with high perturbation. Assign low numbers

TABLE 1.
 $F0$ and Formant Values, in Hz

Parameter	Male	Female
$F0$	120	240
$F1$	700	900
$F2$	1250	1450
$F3$	2500	2500
$F4$	3500	3500
$F5$	4000	4000

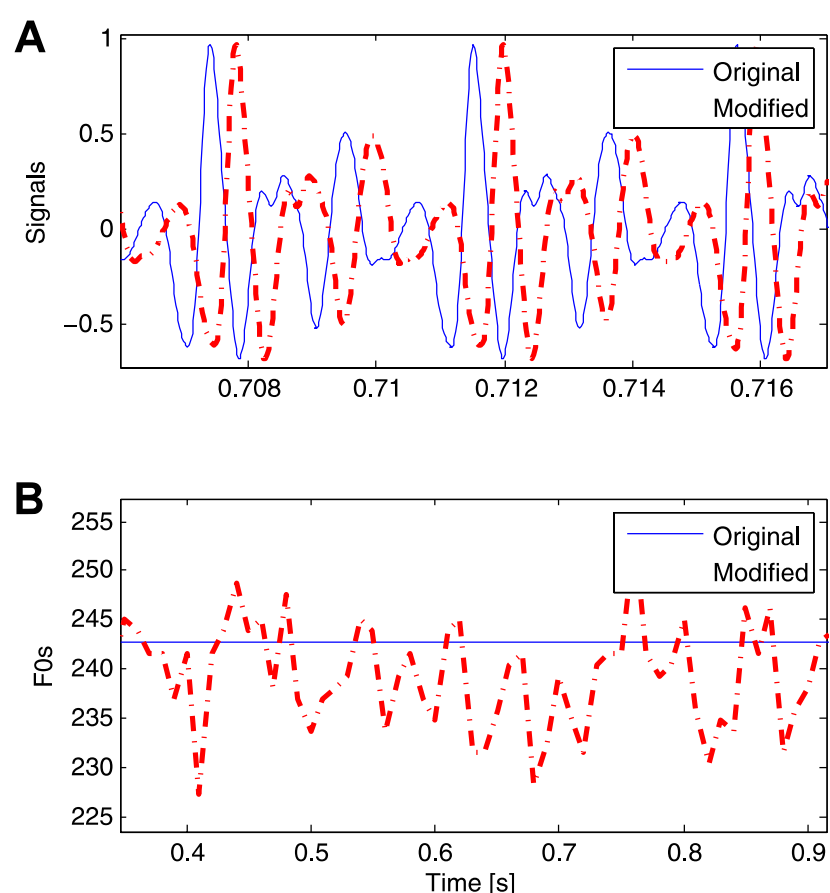


FIGURE 1. Stimuli 0 and 7 (A) waveform and (B) F_0 contours.

to stimuli with low perturbation. Before you start, you may hear stimuli close to both ends of the scale and one intermediate stimulus to be familiar with the stimuli set as many times you may need.”

By performing this orientation, we verified that subjects understood the instructions and were capable of doing the test. After the training, all subjects were asked for a descriptive name for the perturbation they heard.

Intramodal matching procedure. For test 2, we used the intramodal matching method introduced in this article as a procedure to evaluate roughness (intramodal matching procedure [IMP]). Seven stimuli—as indicated in Table 2 (from 1 to 7)—were presented four times in a random order for each listener for a total of 28 trials to constitute the patient stimuli. A graphical software interface with seven buttons that reproduce seven reference stimuli and a patient button that reproduces the 28 trials individually was designed to perform the matching. Instructions to subjects were “when you press the patient button, you will hear the vowel /a/ with a degree of perturbation; your task is to match the patient stimulus with stimuli available on the screen through seven buttons that you can press anytime to find the best match. You can also press the patient button as many times you may need until you decide the match, then you must write the button number in the empty box and press the next button to hear a new patient stimulus and repeat the procedure.” Each stimulus acts four times as a patient voice to be compared against the seven stimuli of the reference set.

Tests 1 and 2 were presented with headphones in a silent room in a balanced order. Each fundamental frequency for each test was also presented in a balanced order.

Subjects

Expert listeners group. Ten speech therapists with >20 years of experience of working in speech evaluation, in

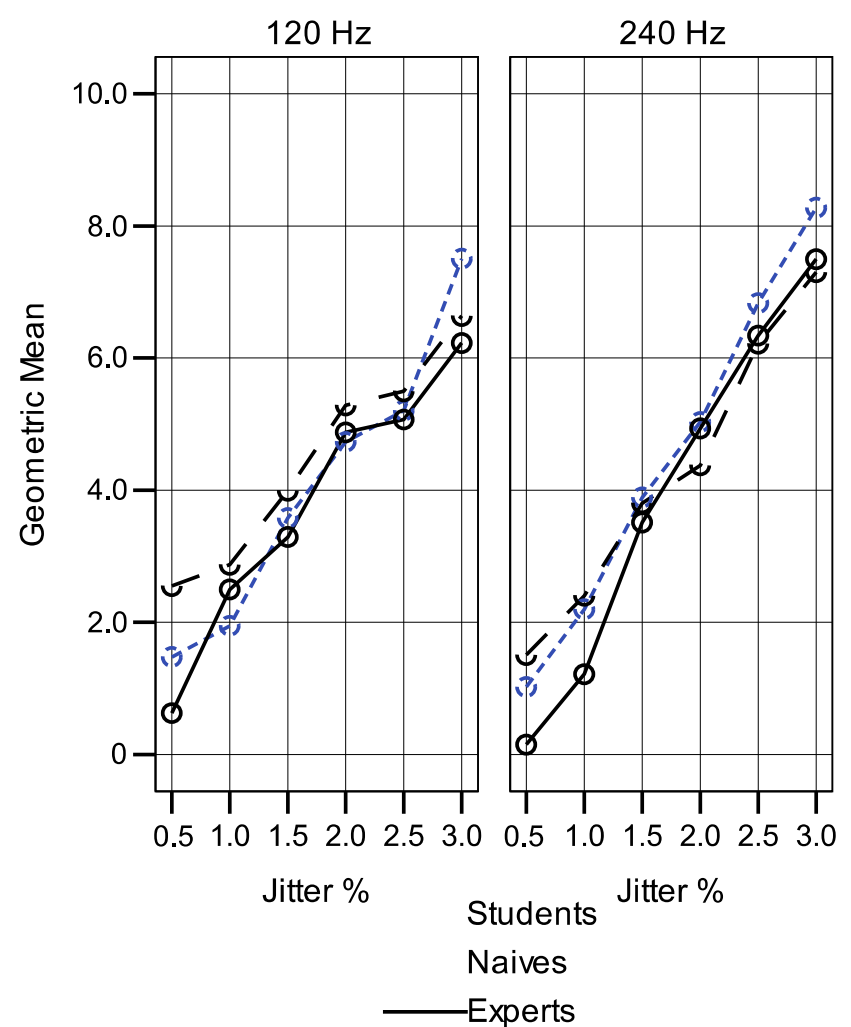


FIGURE 2. Geometric means of responses for ME-CL.

public hospitals, private clinics, and professional voice formation schools of Buenos Aires, participated in the experiments as the expert group (mean age, 54.5 years; standard deviation [SD], 11.8). They usually work with professional voice patients such as singers, speakers, teachers, and actors.

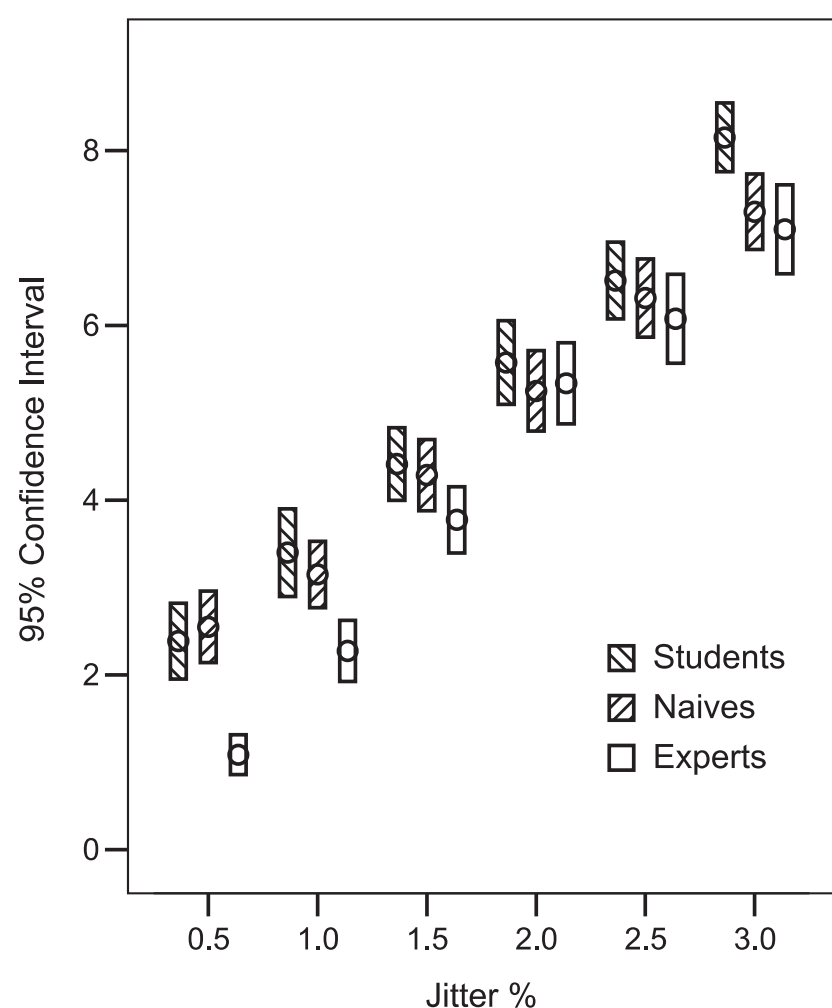


FIGURE 3. Confidence intervals for the ME-CL method.

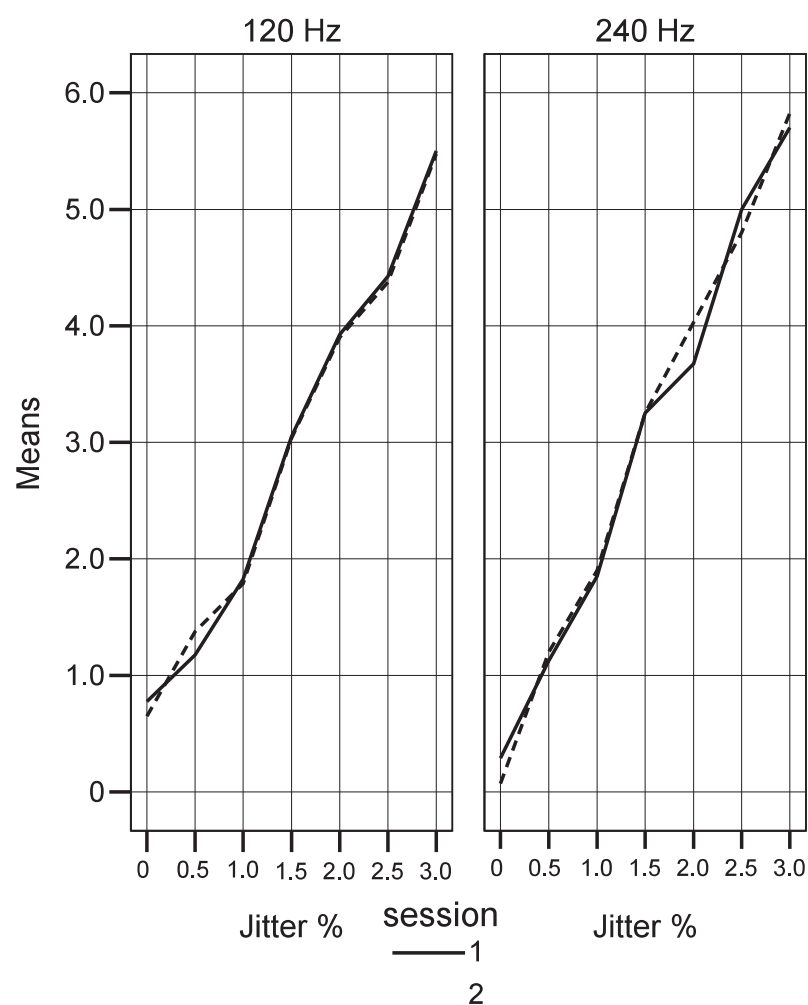


FIGURE 4. Means of experts for IMP for both sessions.

Student listeners group. Ten students from the third and fourth year of the speech therapist program at the University of Buenos Aires functioned as the student group (mean age, 23.2 years; SD, 0.91).

Students have theoretical formation about audio-perceptual scales with no clinical experience at all.

Naïve listeners group. Ten subjects without any theoretical or practical experience listening to vocal speech alterations participated as the naïve group of listeners (mean age, 36.7 years; SD, 17.01).

All subject groups participated in test 1. Both expert and naïve listeners participated in a second session 2–3 weeks later. These two groups of subjects participated in test 2. They also participated in a second session 2–4 weeks after the first session for test-retest control. The *SPSS 13* (SPSS Inc, Chicago, Illinois) package was used to perform statistical analysis.

RESULTS

Estimation experiments

Geometric means of subject responses to ME-CL are presented in Figure 2. As expected, most subjects assigned high perturbation values for high jitter stimuli. Average responses of all groups could fit close to a linear function. Naïve subjects present the smallest sensation range. All listeners show less discrimination to stimuli with low F_0 . At this low frequency voice, listeners show confusion for stimuli 5 (2.00%) and 6 (2.5%). Figure 3 presents confidence intervals for the three groups of listeners. Up to jitter values of around 1.5%, experts show better confidence intervals and less standard deviations $SD \leq 1.7$ (not shown). For high jitter values, dispersion of responses is similar.

After listening to orientation stimuli at the beginning of the experiments, the subjects' sensations were reported as roughness (53%), vowel distortion (30%), and other similar responses (17%). Test duration was 3 minutes.

Matching experiments for experts

Matching functions are known to be a linear combination of two magnitude estimation functions which in turn are considered to be assigned with numbers.² In the present article, we used intramodality matching using exactly the same continuum, so the matching function should be similar to the estimation ME-CL ones. Matching functions are shown in Figure 4 confirming the previously mentioned statement.

In addition, experts produced almost identical results for both sessions. Once again, subjects, as expected, assigned high perturbation values to stimuli with high jitter. Independent of task order, subjects reported that the matching experiment was an easy task to perform. Average experiment duration was 5 minutes.

Statistical analysis using the general linear model tests of between-subject effects for stimulus estimation in ME-CL is ($df = 5$; $F = 189.51$; $P = 0.00$) and for matching in IMP is ($df = 6$; $F = 1212.45$; $P = 0.00$). Analysis of variance of subject responses for methods ME-CL and IMP is also significantly different ($df = 1$; 307.48 ; $P = 0.00$).

TABLE 3.
Intra-Expert and Interexpert Agreement and Reliability Between Sessions 1 and 2

Parameter	120 Hz (\pm SD)	240 Hz (\pm SD)
Intra-expert		
Exact agreement	39.29<54.85%<67.86 (\pm .13)	35.71<76.28%<89.29 (\pm 0.18)
Mean evaluations \pm 1 scale value	100<93.88%<75 (\pm 0.06)	100<98.47%<89.29 (\pm 0.03)
Mean Pearson r	0.771<0.898<0.965 (\pm .06)	0.887<0.959<0.988 (\pm 0.03)
Mean Spearman ρ	0.754<0.896<0.966 (\pm .07)	0.866<0.963<0.992 (\pm 0.03)
Interexpert		
Exact agreement	53.57%	72.86%
Mean evaluations \pm 1 scale value	93.21%	97.86%
ICC (2,1)	0.834	0.929

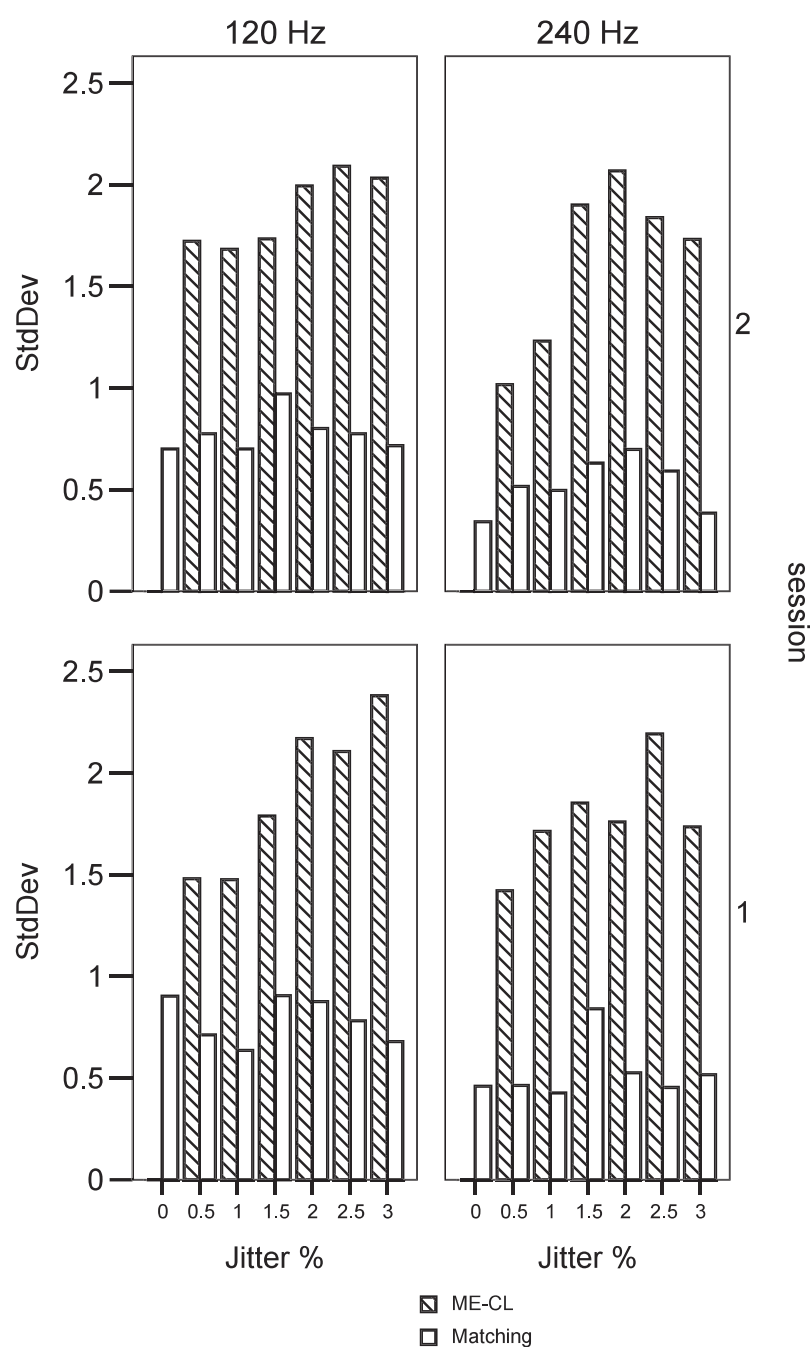


FIGURE 5. Standard deviation of expert responses to ME-CL and IMP.

Test-retest matching experiments for expert listeners

Experts who participated in matching experiments IMP participated in a second session 2–4 weeks later, making it possible to calculate reliability estimates (Figure 4). Both intra-agreement and interagreement and reliability were measured by exact agreement, ± 1 scale value, Pearson moment r , Spearman ρ , and intraclass correlation coefficients ICCs (2,1) for single expert comparisons (Table 3).

Comparison of methods for expert listeners

The ease of the task referred by listeners was verified by comparing the two tasks using a measurement of dispersion. Figure 5 confirms that impression by comparison of standard deviation of responses gathered from the two methods. Taking deviations to matching responses as a reference, deviations for ME-CL responses for 120 Hz are around double; for 240 Hz, the difference rises to triple.

As the main objective of this article was to compare the presented methods, we calculated the expert group entire coherence when performing both methods. Overall, expert average responses for each jitter level from one session to another was calculated for ME-CL and IMP by the ICC (2,10). Results are presented in Table 4. Analyzed as the coherence of the entire group, average results of both F_0 s show ICC values of 0.84 for ME-CL and 0.94 for IMP both with $P = 0.00$.

Matching experiments for naïve listeners

According to ME-CL results, naïve listeners produce responses with the highest standard deviation on average and the worst stimuli discrimination. Related to performance, they appear to be in one extreme, being the expert listeners in the other

TABLE 4. ICCs, Confidence Intervals, and F Test for Expert Raters for Two Sessions of Both Estimation and Matching Tests for All Stimuli, for 120 Hz Stimuli and 240 Hz Stimuli

Parameter	ICC	95% Confidence Interval	F Test With True Value 0
Estimation			
120 Hz			
Single	0.64*	$0.55 < \rho < 0.72$	$(F = 4.59; df = 190; \rho = 0.00)$
Mean	0.78	$0.71 < \rho < 0.84$	$(F = 4.59; df = 190; \rho = 0.00)$
240 Hz			
Single	0.80*	$0.73 < \rho < 0.85$	$(F = 9.81; df = 191; \rho = 0.00)$
Mean	0.89	$0.84 < \rho < 0.92$	$(F = 9.81; df = 191; \rho = 0.00)$
Matching			
120 Hz			
Single	0.83*	$0.79 < \rho < 0.87$	$(F = 11.02; df = 223; \rho = 0.00)$
Mean	0.91	$0.88 < \rho < 0.93$	$(F = 11.02; df = 223; \rho = 0.00)$
240 Hz			
Single	0.93*	$0.91 < \rho < 0.95$	$(F = 27.08; df = 223; \rho = 0.00)$
Mean	0.96	$0.95 < \rho < 0.97$	$(F = 27.08; df = 223; \rho = 0.00)$

Two-way random-effects model. Type A ICC using an absolute agreement definition.

* The estimator is the same, whether the interaction effect is present or not.

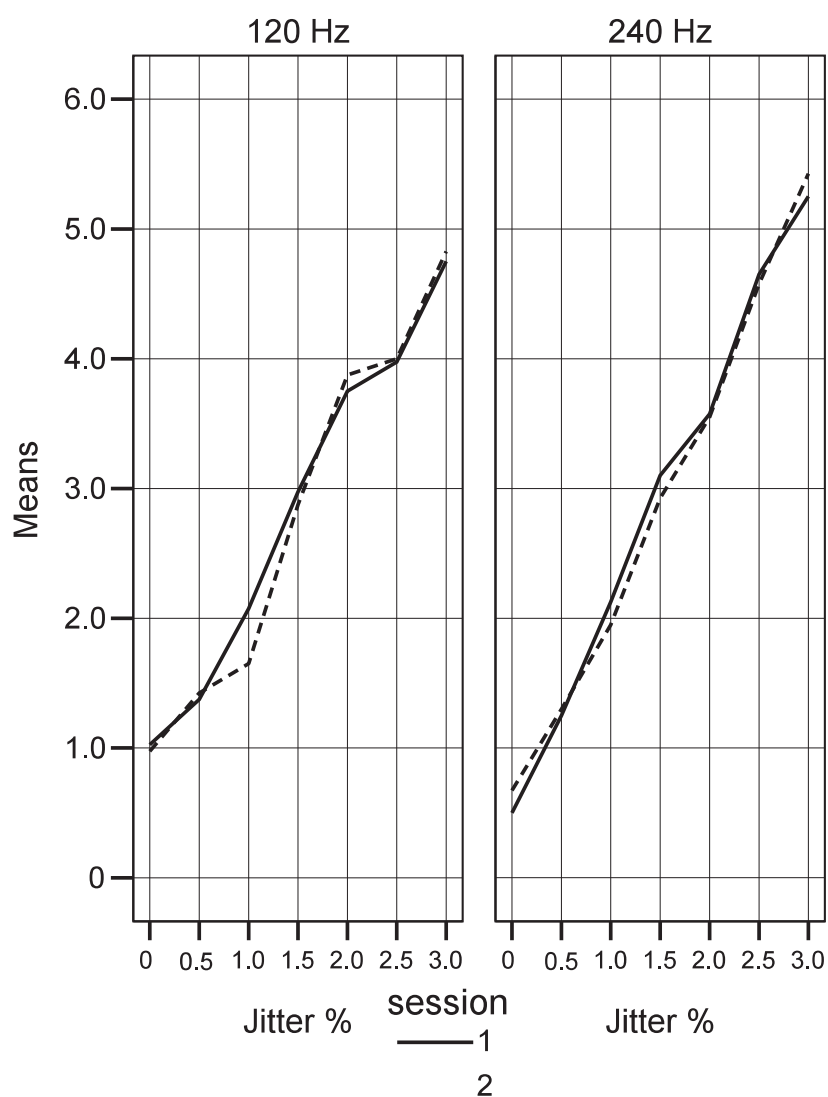


FIGURE 6. Means of naïve responses for IMP for both sessions.

extreme with students somewhere in the middle. At this point, it is interesting to analyze naïve listeners' performance to matching.

Test-retest matching experiments for naïve listeners

Perturbation mean responses for naïve listeners using matching are presented in Figure 6. These two functions appear to be a carbon copy of the functions obtained by the experts. Naïve subjects reported that the matching experiment was an easier task to perform than estimation. Naïve listeners could be seen as completely new speech therapists without experience. For this reason, the computation of ICC using naïve listeners could be an indicator that the matching method is suitable for nonexperts as well. We presented the matching

method twice to ten naïve listeners and the results are summarized in Table 5.

Comparison of methods for naïve listeners

The ease of the task referred by naïve listeners was again verified by comparison of the two tasks using a measurement of dispersion. In Figure 7, more than double deviation is observed for ME-CL responses when compared with IMP responses.

Statistical analysis tests of between-subject effects for ME-CL are ($df = 5$; $F = 74.22$; $P = 0.00$) and for matching ($df = 6$; $F = 452.36$; $P = 0.00$). Here, obtained functions are similar to those obtained for experts. Analysis of variance of subject responses for ME-CL and IMP is ($df = 1$; $F = 343.20$; $P = 0.00$).

DISCUSSION

Results of ME-CL experiments show that low jitter values of 0.5% and 1% are not clearly discriminated precisely because listeners are known to have poor discrimination. Experts perform better than inexperienced listeners for these stimuli. For high jitter values, inexperienced listeners present similar deviation like before, and experts increase the dispersion of their responses. In general, subjects show a smaller range for the 120 Hz stimuli than the observed for 240 Hz. It is not clear why this happens, but the shape of the 120 Hz functions indicates that subjects may have a tendency to use a category scale of few points. Meanwhile, responses to 240 Hz present a good expansion of the numeric scale.

Results of IMP experiments show similar functions as obtained for ME-CL. Test-retest evaluation for experts and naïve listeners produced almost identical functions in both sessions. The main result is that subjects clearly reduce the variability of their responses using IMP. Standard deviations for the IMP method are significantly lower than those obtained for the ME-CL method.

Test and retest experiments for expert listeners show a better score of the ICC parameter for IMP than for ME-CL as summarized in Table 4.

Under IMP, naïve listeners behave more like expert listeners according to the comparison of deviations between both methods. They also show a high score for the ICC parameter

TABLE 5.
ICCs, Confidence Intervals, and F Test for Naïve Raters for Two Sessions of Matching Tests for Naïve Listeners

Parameter	ICC	95% Confidence Interval	F Test with True Value 0
120 Hz			
Single	0.75*	$0.69 < \rho < 0.80$	$F = 6.95$; $df = 224$; $\rho = 0.00$
Mean	0.86	$0.81 < \rho < 0.89$	$F = 6.95$; $df = 224$; $\rho = 0.00$
240 Hz			
Single	0.88*	$0.84 < \rho < 0.91$	$F = 15.32$; $df = 224$; $\rho = 0.00$
Mean	0.94	$0.92 < \rho < 0.95$	$F = 15.32$; $df = 224$; $\rho = 0.00$

Two-way (both people and measure effects) is a random-effect model. Type A ICC using an absolute agreement definition.

* The estimator is the same, whether the interaction effect is present or not.

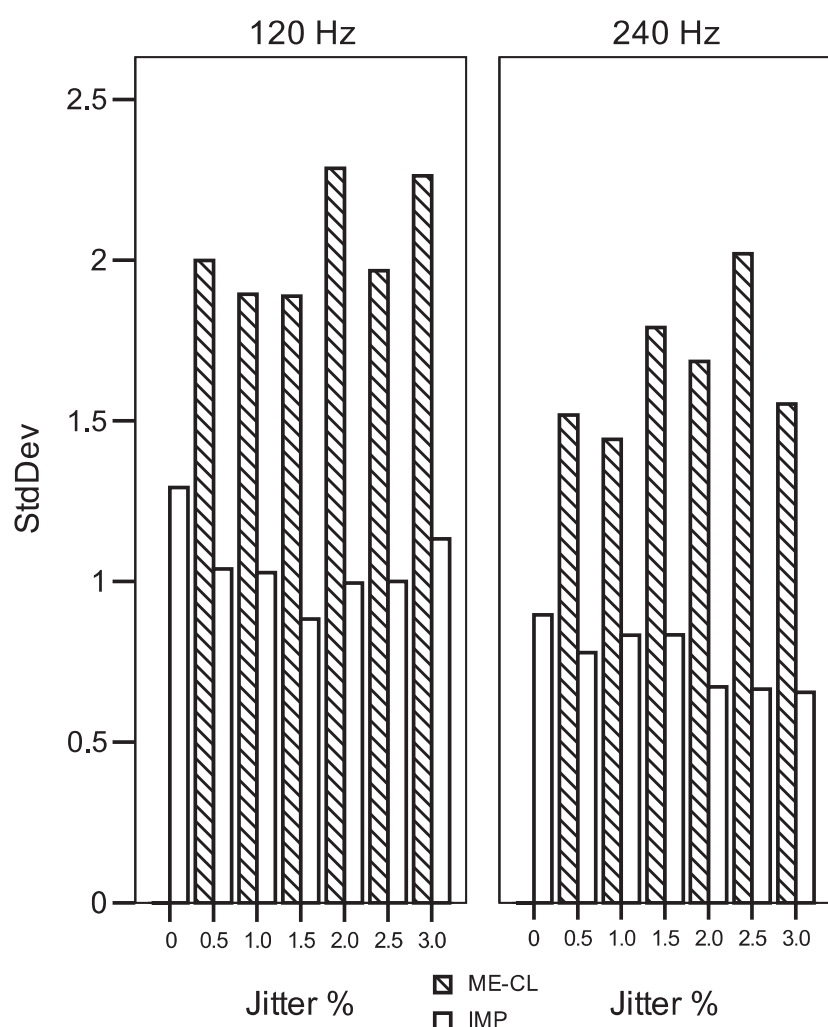


FIGURE 7. Standard deviation of naïve responses to ME-CL and IMP.

but not at the same level of expert scores for matching. Nevertheless, naïve listeners show better ICC than ICC values of expert listeners tested with ME-CL. Using matching, both expert and naïve listeners improved discrimination at low jitter values. Compare jitter values of Figures 5 and 6 against same jitter values of Figure 2.

CONCLUSIONS

We conclude that the consequence of presenting a defined set of external references under an iterative matching procedure with fixed references effectively favors intra-agreement and inter-agreement between expert responses.

Future steps are to repeat this matching procedure varying harmonics-to-noise relation. Experiments for clinical use are planned using a software tool to synthesize a set of external references. In that scenario, the speech therapist could choose both a reference F_0 and overall intensity, according to patient's sustained vowel production, and reproduce external physical references to be matched with patient voice.

Acknowledgments

The authors are grateful to the National Research Council of Argentina CONICET for their financial support. The authors are also very much in debt to the anonymous reviewers, for their language revisions and clarity improvements, and Dr. Guirao, who provided us valuable psychoacoustic details of her designs.

REFERENCES

- Kreiman J, Gerratt BR, Ito M. What and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am*. 2007;122:2354–2364.
- Kempster GB, Abbot KV, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
- Nawka T, Anders LC, Wendler J. Die auditive Beurteilung heiserer Stimmen nach dem RBH-system. *Sprache—Stimme—Gehör*. 1994;18:130–133.
- Hirano M. *Clinical Examination of the Voice*. New York: Springer; 1981.
- Guirao M. A single scale based on ratio and partition estimates. In: Gescheider, Bolanowski, eds. *Ratio Scaling of Psychological Magnitudes*. New Jersey: L. Erlbaum Assoc; 1991:59–78.
- Eisler H, Guirao M. Conventional magnitude estimation versus converging limits procedures: a comparison. In: Preis Anna, ed. *Fechner Day 97, Proceedings XIIth Annual Meeting of the Internat. Soc. for Psychophysics*. Poznam: Polonia; 1997:149–154.
- Sofranko JL, Prosek RA. The effects of levels and types of experience on judgment of synthesized voice quality. *J Voice*. 2014;28:24–35.
- Kreiman J, Gerratt BR. Perception of aperiodicity in pathological voice. *J Acoust Soc Am*. 2005;117:2201–2211.
- Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial and frame work for future research. *J Speech Hear Res*. 1993;36:21–40.
- Gerratt BR, Kreiman J. Measuring voice quality with speech synthesis. *J Acoust Soc Am*. 2001;110:2560–2566.
- Granqvist S. The visual sort and rate method of perceptual evaluation in listening tests. *Logoped Phoniatr Vocol*. 2003;28:109–116.
- Gould J, Waugh J, Carding P, Drinnan M. A new voice rating tool for clinical practice. *J Voice*. 2011;26:163–170.
- Patel SA, Shrivastav R, Eddins DA. Perceptual distances of breathy voice quality: a comparison of psychophysical methods. *J Voice*. 2011;24:168–177.
- Patel SA, Shrivastav R, Eddins DA. Developing a single reference signal for matching breathy voice quality. *J Speech Lang Hear Res*. 2012;55:639–647.
- Stevens SS. *Psychophysics*. New York: Wiley; 1975.
- Guirao M, Garavilla JM. Perceived roughness of amplitude-modulated tones and noise. *J Acoust Soc Am*. 1977;60:1335–1338.
- Fastl H, Zwicker E. Roughness. In: *Psychoacoustics*. Berlin, Heidelberg: Springer; 2007:257–264.
- Bonilha HS, Deliyski DD. Period and glottal width irregularities in vocally normal speakers. *J Voice*. 2007;22:699–708.
- Borzzone de Manrique AM. *Manual de Fonética Acústica*. Buenos Aires: Hachette; 1980.
- Gurlekian JA, Elisei NY, Eleta M. Caracterización articulatória de los sonidos vocálicos del español de Buenos Aires mediante técnicas de resonancia magnética [In Spanish]. *Fonoaudiológica*. 2004;50(2):7–14.
- Ruinskiy D, Lavner Y. Stochastic models of pitch jitter and amplitude shimmer for voice modification. In Proc. of 25th IEEE Convention of Electrical and Electronics Engineers. Dec. Israel. 2008.
- Patel SA, Shrivastav R, Eddins DA. Identifying a comparison for matching roughness. *J Speech Lang Hear Res*. 2012;55:1407–1422.
- Alzamendi GA, Schlotthauer G, Rufiner HL, Torres ME. Evaluation of a new model for vowels synthesis with perturbations in acoustic parameters. *Latin Am Appl Res*. 2013;43:225–230.
- Titze IR. *Workshop on Acoustic Voice Analysis: Summary Statement*. Technical report. Denver, USA: National Center for Voice and Speech; 1995.
- Torres ME, Schlotthauer G, Rufiner HL, Jackson-Menaldi MC. Empirical mode decomposition. Spectral properties in normal and pathological voices. In: Vander Sloten J, Verdonck P, Nyssen M, Haueisen J, eds. *ECIFMBE 2008, IFMBE Proceedings*. Vol. 22. Berlin, Heidelberg: Springer-Verlag; 2009:252–255.
- Rabiner LR, Shafer RW. *Digital Processing of Speech Signals*. Boston: Pearson Education; 2006.

APPENDIX**Original instructions for Magnitude Estimation
Converging Limits method (ME-CL)**

“Usted oirá la vocal /a/, su tarea consiste en evaluar el grado de perturbación dando a cada estímulo un número de acuerdo a una escala numérica de su preferencia. Siempre asigne números más grandes a los estímulos con mayor perturbación. Asigne números más bajos a los estímulos con menor perturbación. Antes de comenzar y para familiarizarse con el conjunto de estímulos usted podrá oír los estímulos cercanos a los extremos y también un estímulo intermedio, tantas veces como lo crea necesario.”

**Original instructions for Intramodal Matching
Procedure (IMP)**

“Cuando usted oprima el botón que dice “Paciente,” escuchará una vocal/a/con un grado de perturbación; su tarea consiste en comparar el estímulo “Paciente” con los estímulos disponibles en la pantalla a través de siete botones que podrá oprimir en cualquier momento para encontrar la mejor igualdad. Usted puede también oprimir el botón Paciente todas las veces que lo requiera hasta que decida la igualdad, entonces escriba el número del botón elegido en la ventana vacía y oprima el botón “Próximo” para escuchar un nuevo estímulo “Paciente” y repetir el procedimiento.”