How Reliable Is the Auditory-Perceptual Evaluation of Phonation Onset Hardness?

*,‡,§,||,¶,**Youri Maryn, and †Sofie Poncelet, *Wilrijk/Antwerp, †Molenbeek, ‡§Ghent, ||Louvain-la-Neuve, ¶Antwerp, and **Lokeren, Belgium

Summary: Background. Generally, phonation onset hardness has been similarly defined and categorized throughout literature. There are typically three categories: soft, breathy, and hard onset. Phonation onset hardness is relevant in both assessment and treatment of various voice disorders and is usually determined on an auditory-perceptual basis. However, as far as we know, reliability of clinical auditory-perceptual indications of phonation onset hardness has not been investigated yet. Therefore, this study aimed at evaluating intrarater as well as interrater reliability of auditory-perceptual assessment of phonation onset hardness by speech and language pathologists.

Methods. After deidentification, the sentence-initial word [e.rst] was extracted from read text recordings of 20 subjects with various voice disorders and complaints. These 20 samples were purposefully selected to represent as much as possible the whole phonation onset hardness continuum. The auditory-perceptual rating protocol consisted of the paired comparison paradigm, in which all listeners were asked to perceptually compare phonation onset hardness of every fragment with the other 19 samples. This resulted in a ranking of the 20 samples, from hardest to softest phonation onset. Four speech and language pathologists agreed to compare phonation onset hardness according to this paradigm. The single-measures intraclass correlation coefficient (ICC) for absolute agreement was applied to determine the degree of reliability within as well as between raters.

Results. Intrarater ICC's showed acceptable reliability for two raters, yet poor reliability for the other two raters. Interrater ICC's demonstrated low reliability in general. Zooming in on the ratings shows that three of four evaluators agreed on which sample had the softest phonation onset, and that only two raters agreed on which recording had the hardest phonation onset.

Discussion. It can be concluded that there is considerable variability within and across raters when asked to indicate the word with the hardest phonation onset. This is the first study to approach auditory interpretation of phonation onset hardness with the paired comparison task. Questions are raised about the clinical utility of phonation onset hardness perception, the potential role of training, and the importance of a more objective yet clinically feasible measurement tool.

Key Words: Phonation onset—Hardness—Auditory-perceptual ratings—Paired comparison—Sound intensity level.

INTRODUCTION

Across literature, phonation onset has been similarly defined and classified in the three following categories: soft, breathy, and hard onset.[1,2,3,4,5] Soft phonation onset (SPO; ie, simultaneous or gentle onset) is characterized by the simultaneous activity of exhaling and adduction of the vocal folds. During this prephonatory period, the posterior cricoarytenoid muscle has been described to relax while tension in the lateral cricoarytenoid muscle gradually increases. Vocal rise time[1] for gentle onset has been found to vary between 163 and 325 ms[2] and 60-120 ms.[4] Breathy phonation onset (ie, aspirated onset) occurs when there is exhalation and transglottal airflow before vocal fold adduction and thus prior to initiation of phonation. During this onset type vocal folds approach medial compression but are not entirely adducted. The posterior cricoarytenoid muscle is active across this prephonatory period, while tension in the lateral cricoarytenoid muscle increases gradually. Vocal rise time for breathy onset has been described to vary between 65 and 243 ms[2] and 100-190 ms.[4] Hard phonation onset (HPO; ie, abrupt onset or glottal attack) is characterized by the abrupt medial compression of vocal folds before exhalation is initiated and thus prior to transglottal airstream. The posterior cricoarytenoid muscle's activity is decreased whereas the lateral cricoarytenoid muscle is increased well before phonation onset, which results in prephonatory high subglottal pressure. Vocal rise time for hard onset has been found to differ between 18 and 50 ms[2] and 18-42 ms.[4] Clinically, SPO is considered the least phonotraumatic and thus optimal initiation of phonation, in contrast to HPO.

Several studies have been investigating the frequency of phonation onset hardness/abruptness types (and particularly the attacked phonation onset) in people with and without a voice disorder. From Andrade et al[6] it can be learned that patients with muscle tension dysphonia as well as benign vocal folds lesions (both unilateral and bilateral, but

Accepted for publication April 1, 2020.
Level of Evidence: Level 2b: cohort study of rater reliability.
From the *European Institute for ORL-HNS, Department of Otorhinolaryngology and Head & Neck Surgery, GZA Sint-Augustinus, Wilrijk/Antwerp, Belgium; †Multidisciplinary clinics 't Ketje and 't Zinneke, Molenbeek, Belgium; ‡Department of Speech, Language and Hearing Sciences, University of Ghent, Ghent, Belgium; §Faculty of Education, Health and Social Work, University College Ghent, Ghent, Belgium; ||Faculty of Psychology and Educational Sciences, Ecole de Logopédie, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; ¶Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; and the **Phonanium, Lokeren, Belgium.

Address correspondence and reprint requests to Youri Maryn, European Institute for ORL-HNS, Department of Otorhinolaryngology and Head & Neck Surgery, GZA Sint-Augustinus, Wilrijk/Antwerp, Belgium. E-mail: youri.maryn@gza.be
Journal of Voice, Vol. ■■, No. ■■, pp. ■■–■■
0892-1997
© 2020 The Voice Foundation. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jvoice.2020.04.006

[1]Vocal rise time is the duration from the onset of sound to the moment at which the amplitude of the acoustic sound pressure wave reaches a steady state[4].

especially bilateral vocal fold nodules) demonstrated higher frequencies of HPO than control subjects with vocal fold pathology. Uygun et al[7] also found significantly higher HPO frequency in patients with vocal fold nodules or muscle tension dysphonia than in participants without a voice disorder. However, are abrupt phonation onsets cause of a voice disorder, or induces a voice disorder use of abrupt phonation onsets? The actual cause-effect mechanisms behind HPO and voice disorders still remains unclear and require further investigation. Regardless of HPO predisposing/precipitating or perpetuating a voice disorders, taking phonation onset type into account is pertinent in clinical assessment and treatment of some voice disorders, such as vocal process granuloma[8,9] and functional voice disorders.[10]

Several objective/quantitative methods have been developed to objectively approach phonation onset, such as vocal rise time,[2,11] vocal attack time,[5,12,13,14] voice onset coordination,[15] onset oscillatory time[16,17] or first peak of acoustic waveform derivative.[18] However, as far as we know, none of these methods have been implemented in clinical voice assessment protocols yet, probably because the majority of clinicians lack the required (a) signal acquisition systems, for example, for high-speed videolaryngoscopy, electroglottography, and/or airflow signal recording, and/or (b) software systems for (simultaneous) analysis of these signals. Until such methods become clinically feasible, auditory-perceptual evaluation is hypothesized to stay primary method in clinical evaluation of phonation onset abruptness/hardness.[18]

For auditory-perceptual judgments concerning voice quality to be clinically meaningful, it is essential for differences within raters and between raters to be as minimal as possible. Reliability is key when it comes to clinical usefulness of subjective auditory-perceptual judgments of any kind of health-related aspect. In the past, interrater and intrarater reliability have been found to be variable for existing protocols for voice quality judgments, even when judged by experienced listeners.[19] During the last decades, for example, voice quality rated with the GRBAS scales[20] has been subjected to numerous reliability assessments. A study of Yamaguchi et al[21] comparing the use of GRBAS by Japanese and American raters, showed that for both cultural groups G, B, and R were judged more reliable than A and S. The judgments of G, B, and R were less variable than judgments for A and S. Eadie and Baylor[22] investigated the intra- and interrater reliability of perceptual judgment of dysphonia severity, roughness, and breathiness by inexperienced listeners using a visual analog scale before and after training. They found that both intra- and interrater reliability improved after training. See, for example, Roy et al[23] or Barsties and De Bodt[24] for more profound discussion on auditory-perceptual evaluation of voice signals. However, as far as we know, reliability assessment of phonation onset abruptness is scarce and has only been conducted in one study.[25] Four healthy male speech-language pathology students were trained in and recorded while producing three isolated vowels with seven degrees of phonation onset abruptness while keeping maximum sound level and stimulus duration as constant as possible. Six speech-language pathologists and five phoneticians, all professionally active yet untrained in judging phonation onset, were asked to rate onset of all stimuli on a 7-point scale ranging from "extremely soft" to "extremely hard." With an interrater intraclass correlation coefficient (ICC) of 0.74, it was concluded that the raters were fairly reliable in evaluating phonation onset abruptness.[25] Although this well-designed study controlled for vowel type and degree of phonation onset hardness, the generalizability of its results can be questioned because it only contained isolated vowels by only four speakers mimicking different hardness levels.

Literature unanimously relates HPO and voice disorders. Consequently, phonation onset is considered as one of the legit voice aspects to be evaluated during clinical voice assessment. However, not much is known about the reliability of phonation onset evaluations in a clinically representative set of voice recordings. Therefore, the purpose of the current study was to examine the intrarater as well as interrater reliability of phonation onset abruptness ratings by speech-language pathologists without specific training of perceptual assessment of phonation onset evaluation.

METHODS

Sample selection

To obtain a set of 20 audio fragments representing the phonation onset continuum (ie, from clearly aspirated to clearly hard), audio recordings in a database of the first authors (YM) with 405 voice samples were retrospectively scanned on phonation onset by the second author (SP; who followed a postgraduate specialty course on voice disorders). To have a homogeneous set of phonation onsets, samples of males and children were a priori excluded. Only female voice samples were carefully listened to and subjectively assigned into one of the three phonation onset categories. This procedure finally resulted in a convenience sample of 20 recordings that were purposefully selected to represent as much as possible the continuum from clearly aspirated phonation onset to clearly hard/attacked phonation onset.

This scanning and selection by SP included listening to the vowel-initial word "Eerst" (or phonetically [e.rst]) at the start of the third sentence of the phonetically balanced Dutch text "Papa en Marloes."[26] When found to be appropriate for completion of the phonation onset continuum, this fragment was extracted and saved as a separate WAV-file. This was all done in the program Praat (Paul Boersma and David Weenink, Institute of Phonetic sciences, University of Amsterdam, The Netherlands) by a trained speech-language pathologist (SP) enrolled in a postgraduate course on voice disorders. All voice samples were deidentified prior to evaluation.

At the end of this procedure, there was a batch of 20 "Eerst" samples with a variety in phonation onset abruptness. These samples stemmed from 20 Dutch-speaking females (mean age: 57 years; age range: 23-87 years). Because of voice-related complaints, they underwent standard clinical voice assessment and recording at European Institute for

ARTICLE IN PRESS

TABLE 1.
Laryngostroboscopic Diagnoses

Diagnosis	Number ($N = 20$)
Hypotrophy	3
Voice nodules	2
Cyst	2
Edema	1
MTD	1
Hemorrhage	1
Laryngitis	1
Granuloma	1
Bilateral vocal fold paralysis	1
Unilateral vocal fold paralysis	1
No pathology	6

ORL-HNS (Otorhinolaryngology Department, GZA Sint-Augustinus Hospital, Wilrijk, Belgium). Their laryngostroboscopic findings are listed in Table 1.

Auditory-perceptual ratings

Auditory-perceptual assessment was obtained by the paired comparison paradigm. This method to rate degree or severity of phenomena has already proven to be a feasible and reliable method for perceptual voice quality evaluation purposes.[27,28,29,30] In this task, all listeners were asked to perceptually compare phonation onset hardness of every fragment with the other 19 samples. To avoid comparison of two identical samples and double comparison of two identical (however reversed) sample pairs, the number of pairs to be compared (N) was multiplied with N-1 and divided by 2, respectively. This resulted in a total of 190 pairs that were randomly presented for the auditory-perceptual comparisons. Every time a sample was judged to have the HPO, it earned one point. The sample with the SPO acquired zero points. Once all comparisons were made, the points per sample were tallied. The higher the total number of points, the more a voice sample was judged to have HPO relative to the other samples in this study. The result of this paired comparison paradigm was a ranking of the 20 samples, from HPO to SPO. To determine intrarater reliability of auditory-perceptual phonation onset ratings, 19 randomly selected sample pairs (ie, 10% of all pairs) were repeated a second time at the end of this experiment. Consequently, raters were instructed to judge in total 209 pairs on phonation onset hardness.

Raters and instructions

Four Dutch-speaking speech-language pathologists voluntarily agreed to assess phonation onset according to the above-mentioned auditory-perceptual rating procedure. One of them worked as a vocologist, two were enrolled in the postgraduate voice disorders program at the moment of the experiment, and one had no specific experience in the area of voice disorders yet received basic information during her graduate training as speech-language pathologist/audiologist.

Raters were given a brief instruction regarding the rating task. No definition of phonation onset was provided, but all judges stated prior use of rating phonation onset as hard, easy or soft. The raters were instructed to listen to two paired voice samples and then cross the box of the voice sample with the HPO. No other training was offered prior to the experimental ratings of phonation onset hardness. Raters received a printed paper on which all pair numbers were printed. During the experiment, all raters performed the ratings individually in one session. When requested by a listener, a sample could be listened to multiple times.

Sound intensity level

Whether perceived phonation onset hardness is related and has to be attributed to sound intensity level needed to be investigated as well, especially because clear differences in sound intensity level between the 20 samples were observed. The sound intensity levels were determined with the "To Intensity…" command [Minimum pitch = 60 Hz; Time step = 0 s (=auto); Subtract mean = yes] of the program Praat. It consisted of post hoc uncalibrated sound intensity measures in dB. However, because only relative (ie, between-sample differences in) sound intensity measures were of interest to this study, it was decided to continue without calibration toward true sound intensity levels.

Statistics

First, to assess interrater as well as intrarater reliability of the phonation onset hardness ratings in the pairs of samples, the following kind of ICC was calculated, according to the guideline of Koo and Li.[31] The "two-way random effects" ICC model was chosen as we imagined our reliability results to be generalized to any rater with same characteristics as the selected raters in the present study. The "single rater" ICC type was selected because we did not plan to use the mean value of 3 raters for assessment purposes. Because only a dichotomous choice was possible in the paired comparison task, the "absolute agreement" ICC definition was chosen to evaluate whether the four raters chose the same samples to have the hardest onset in the pairs. *IBM SPSS Statistics* for Mac version 20.0 was applied for the calculation of this ICC. ICC is a single value which accounts for degree of proportional correspondence as well as agreement among two or more raters. This index has been shown useful with dichotomous variables. Portney and Watkins[32] suggest that ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability. Second, to assess proportional correspondence between the raters' choices of the hardest within the pairs of phonation onset samples, Cramer's V coefficient of association (ie, V) between two related sets of binary data was calculated. Third, to evaluate difference between the rater's choices, the omnibus Cochran's Q test for $k = 4$ raters followed by post hoc McNemar's nonparametric tests for

related dichotomous data was applied. Fourth, to investigate the proportional relationship between phonation onset hardness expressed as the average ranking across the four raters, and sound intensity level, Pearson product-moment correlation coefficient (ie, r_P) between these two variables.

Ethics

This retrospective study of these phonation onset samples was approved by the Ethics Committee of GZA Sint-Augustinus hospital (approval code 180407BACHEL).

RESULTS

ICC's regarding intrarater reliability in choosing the HPO in pairs of recordings are provided in Table 2. It is good in raters 1 and 2, while poor for raters 3 and 4. For intrarater reliability, 19 pairs were judged twice by each rater. Rater 1 chose in all but one pair the same samples to have the hardest onset. Rater 2 chose differently in two pairs. Raters 3 and 4 judged differently in five and six pairs, respectively. Within-rater reliability in evaluating phonation onset hardness is thus variable.

TABLE 2.
Intrarater Reliability in Auditory-Perceptual Evaluation of Phonation Onset Hardness

	ICC
Rater 1	0.889
Rater 2	0.769
Rater 3	0.392
Rater 4	0.341

Abbreviation: ICC, intraclass correlation coefficient.

Interrater reliability in choosing the HPO in pairs of recordings as well as the similarity between the resultant rankings is given in Table 3. The computed ICC values show that overall interrater reliability is low when using the paired comparison task. When zooming in on the ratings, three out of four evaluators agreed on which sample had the SPO, and only two evaluators agreed on which recording had the HPO. To visually show the interrater variability, a bar graph was drawn with the total number of selections per rater for each voice fragment (Figure 1). Ideally, the four bars are equally high for each individual recording. When

TABLE 3.
Interrater Reliability in Auditory-Perceptual Evaluation of Phonation Onset Hardness

	Rater 2			Rater 3			Rater 4		
	ICC	V	*P*	ICC	V	*P*	ICC	V	*P*
Rater 1	0.145	0.148	0.043	0.245	0.244	0.812	0.460	0.469	0.008
Rater 2				0.301	0.305	0.061	0.197	0.213	0.000
Rater 3							0.410	0.421	0.004

Abbreviations: ICC, intraclass correlation coefficient; *P*, asymptotic significance in McNemar test; V, Cramer's V correlation coefficient.

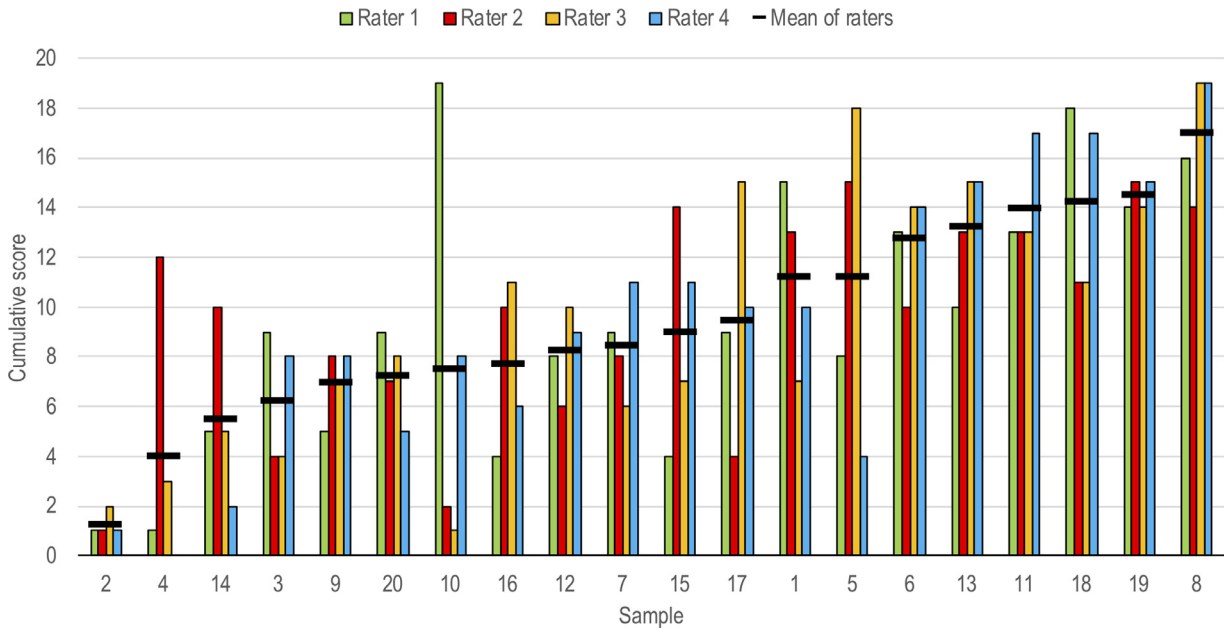


FIGURE 1. Illustration of the limited interrater reliability in auditory-perceptual evaluation of phonation onset hardness: cumulative scores for phonation onset hardness for each of the 20 voice fragments ordered according to the mean ratings.

zooming in on sample 2, all raters agreed this sample to have a very SPO. Sample 19, on the other hand, was quite consistently judged to have very HPO. These two samples may therefore count as representative samples of SPO and HPO, respectively. However, in sample 10, for example, raters were not in agreement at all. Figure 2 shows sound pressure and sound intensity curves of these three exemplary phonation onsets. Additional to the ICC analysis, Cramer's V tests demonstrated low proportional association between the choices (of which of two phonation onsets is harder) of the raters. Furthermore, because Cochran's Q omnibus test revealed significant difference ($Q = 22.955$; $P < 0.001$) between the rater's choices, post hoc McNemar tests were applied and showed significant difference between raters 1 and 2 (chi square = 4.101; $P = 0.043$), raters 1 and 4 (chi square = 6.942; $P = 0.008$), raters 2 and 4 (chi square = 18.278; $P < 0.001$), and raters 3 and 4 (chi square = 8.491; $P = 0.004$).

The scatterplot in Figure 3 illustrates the nonsignificant $r_P = 0.257$ ($P = 0.272$) between mean perceived phonation onset hardness and sound intensity level. This correlation corresponds with $r_P^2 = 0.066$, implying that only 6.6% of the variance in phonation onset abruptness can be accounted for by variance in sound intensity level.

DISCUSSION

Since phonation onset has been related to voice disorders and vice versa, its importance in clinical assessment of voice disorders is undeniable. Because not much is known about the current subjective auditory-perceptual evaluation of phonation onset during voice assessment, this has to be researched in order to make reliable conclusions before taking it into account in voice therapy. In this regard, current results show that intrarater as well as interrater reliability of phonation onset abruptness ratings (ie, choosing which of two samples has the hardest onset) are variable.

Degree of experience with pathological voices could have been an influential factor. Previous research has suggested that even highly experienced listeners differ widely in their judgments about voice quality.[19] In the current study, raters 1 and 2 were vocology trainees enrolled in a voice disorders-specific postgraduate program. Both of them accounted for the highest values of intrarater variability. However, the vocologist who already completed the postgraduate voice disorder program 2 years ago and was supposedly most experienced in clinical management of voice disorders, showed the lowest intrarater reliability. The speech-language pathologist/audiologist not specifically engaged in voice disorders showed the second lowest intrarater

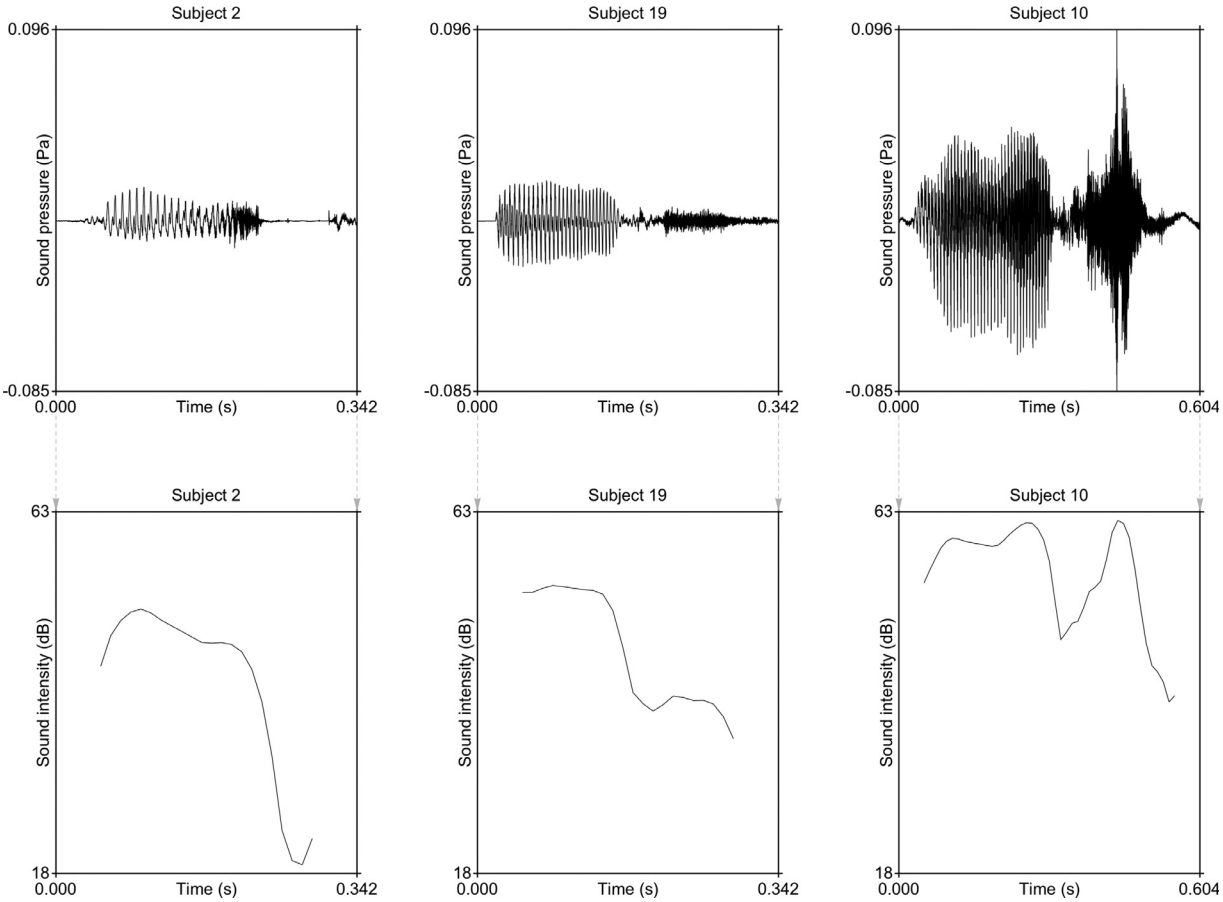


FIGURE 2. Sound pressure and sound intensity curves of three exemplary phonation onsets that were rated in this study. **Left**: phonation onset from subject 2 consistently rated as soft. **Middle**: phonation onset from subject 19 consistently rated as hard. **Right**: phonation onset from subject 10 inconsistently rated in terms of abruptness.

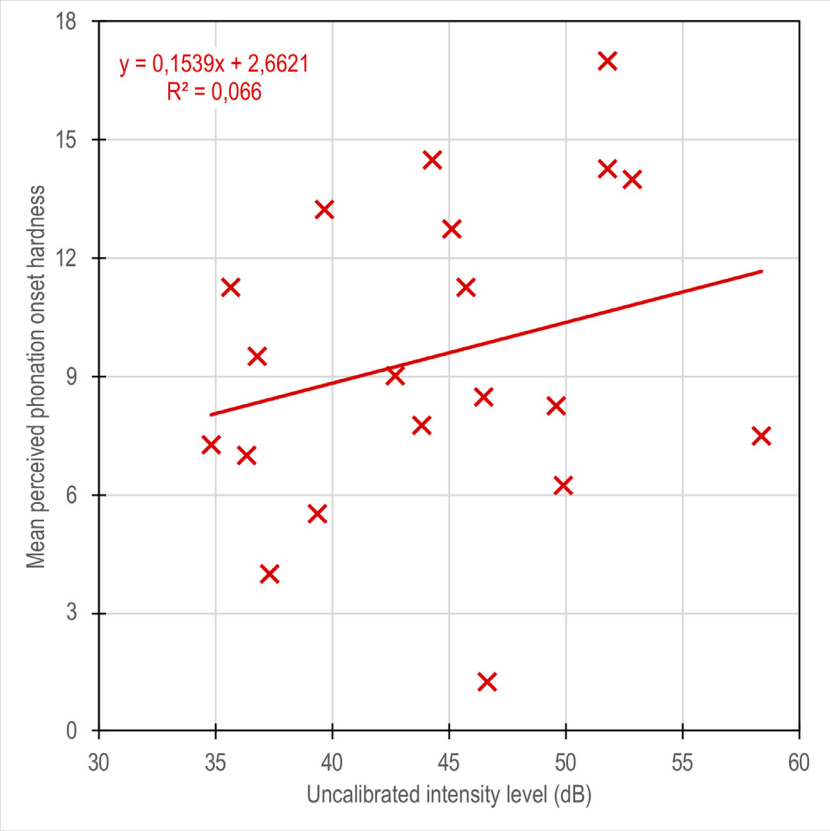


FIGURE 3. Scatterplot to illustrate the proportional relationship between perceived phonation onset hardness and sound intensity level.

reliability. At the moment of experiment, people enrolled in the postgraduate program were actively and profoundly engaged in auditory evaluation of different dimensions of the pathological voice. One of the tasks during this postgraduate program entailed the development a protocol for voice assessment. They were also enrolled in a GRBASI[20] training program which entailed the perceptual scoring of all GRBASI dimensions in comparison to reference voices. It can therefore be assumed that the ongoing auditory "training" they had at that time accounted for their higher intrarater reliability.

Sound intensity level of the recordings may also have played a role, as reported by the raters during the discussion after the experiment. Speech samples were collected retrospectively in the four years preceding the experiment. Although the same sound recording equipment was always used with the microphone always placed similarly relative to the mouth, some of the recordings indeed had higher intensity levels than others, possibly due to the patients' vocal intensity characteristics itself. Figure 2 provides some examples to illustrate the differences in sound intensity level. However, only a very poor and not significant relationship appeared between the average phonation onset hardness ratings and the post hoc sound intensity level measures. Nevertheless, future experiments with phonation onset should consider this and maybe equalize the mean intensity levels of the samples before analyzing them. On the other side, differences in intensity level are expected in any set of clinical voice recordings. Therefore, changing their intensity level before analysis may reduce the clinical representativity of the rating task.

In order to reliably evaluate a specific phenomenon, it is important to have a clear and comprehensive definition of this phenomenon, ie, phonation onset. Throughout literature, phonation onset has been divided into three categories.[4] However, as far as we know, there is no operational definition for the auditory-perceptual evaluation of phonation onset. Furthermore, as Kreiman et al,[19] Yiu et al,[28] and Eadie and Kapsner-Smith[33] already pointed out, use of anchor (ie, reference) recordings increase rater reliability. However, for phonation onset, no set of anchors is currently available to represent the complete continuum from SPO to HPO. Future research may therefore be directed in the development of such a set of recordings.

As classifying phonation onset type or rating its hardness is subjective and clearly accompanied by intrarater as well as interrater variability, adding more objective methods is presumed to improve the clinical phonation onset evaluation protocol. Certain (combinations of) acoustic, high-speed laryngoscopic, aerodynamic, and electroglottographic measures have been developed for that purpose (see Madill et al[18] for an overview of these methods). As far as we know, however, none of these instrumental and less rating-based methods are currently available for clinical use. Therefore, while waiting for such methods to become at clinician's disposal, it is important to be aware of the

variability in phonation onset hardness ratings and thus also to train, for example, speech-language pathology students in this.

To be as much as possible representative for daily voice use and patterns, the stimulus used in this experiment was the word "Eerst" (or phonetically [erst]) extracted from the start of the third sentence of the phonetically balanced text "Papa en Marloes."[26] As far as we know, this was the first time that a phonation onset study used such a habitual voicing context. Previous research in judging phonation onset has been using isolated vowels[25,34,35] or isolated monosyllabic words recorded outside the context of regular phonation.[11] Andrade et al[6] and Uygun et al[7] used respectively custom-made English (the Towne-Heuer text[36]) and Turkish texts in their studies, however only to determine absolute or relative frequency of hard glottal attacks, and thus not to rate degree of phonation onset hardness or type. Furthermore, because vowels [i] and [u] were judged to be initiated less harsh than other vowels,[34] it is also important to include multiple words with different initial vowels in future analyses of phonation onset.

CONCLUSIONS

From current results, it can be concluded that there is considerable variability within and across raters when asked to indicate the HPO. As far as we know, this is the first study to approach auditory interpretation of phonation onset hardness with the paired comparison task. Questions are raised about the clinical utility of phonation onset hardness perception, the potential role of training, and the importance of a more objective yet clinically feasible measurement tool.

REFERENCES

1. Faaborg-Andersen K. Electromyographic investigation of intrinsic laryngeal muscles in humans. *Acta Physiol Scand*. 1957;41:1–149.
2. Koike Y. Experimental studies on vocal attack. *Otorhinolaryngol Clin Kyoto*. 1967;60:663–688.
3. Hirose H, Gay T. Laryngeal control in vocal attack: an electromyographic study. *Folia Phoniatr*. 1973;25:203–213.
4. Behrman A. *Speech and Voice Science*. San Diego, CA: Plural Publishing; 2007:152–153.
5. Orlikoff RF, Deliyski DD, Baken RJ, et al. Validation of a glottographic measure of vocal attack. *J Voice*. 2009;23:164–168.
6. Andrade DF, Heuer R, Hocksteun NE, et al. The frequency of hard glottal attacks in patients with muscle tension dysphonia, unilateral benign masses and bilateral benign masses. *J Voice*. 1999;14:240–246.
7. Uygun MN, Esen Aydinli F, Aksoy S, et al. Turkish standardized reading passage for the evaluation of hard glottal attack occurrence frequency. *J Voice*. 2018;32:51–56.
8. Leonard R, Kendall K. Effects of voice therapy on vocal process granuloma: a phonoscopic approach. *Am J Otolaryngol*. 2005;26:101–107.
9. Patel RR, Pickering J, Stemple J, et al. A case report in changes in phonatory physiology following voice therapy: application of high-speed imaging. *J Voice*. 2012;26:734–741.
10. Chan AK, McCabe P, Madill CJ. The implementation of evidence-based practice in the management of adults with functional voice disorders: a national survey of speech-language pathologists. *Int J Speech Lang Pathol*. 2013;15:334–344.
11. Leeper HA. Voice initiation characteristics of normal children and children with vocal nodules: a preliminary investigation. *J Commun Disord*. 1976;9:83–94.
12. Roark RM, Watson BC, Baken RJ, et al. Measures of vocal attack time for healthy young adults. *J Voice*. 2012;26:12–17.
13. Watson BC, Baken RJ, Roark RM. Effect of voice onset time on voice attack time. *J Voice*. 2016;30:11–14.
14. Baken RJ, Watson BC. Research note: vocal attack time-extended analysis. *J Voice*. 2019;33:258–262.
15. Baillie G, Barone NA, Ludlow CL. *Neural Plasticity in Singers*. Chicago, IL: ASHA Convention; 2013.
16. Patel RR, Walker R, Döllinger M. Oscillatory onset and offset in young vocally healthy adults across various measurement methods. *J Voice*. 2017;31:512.e17–512.e24.
17. Patel RR, Forrest K, Hedges D. Relationship between acoustic voice onset and offset and selected instances of oscillatory onset and offset in young healthy men and women. *J Voice*. 2017;31:389.e9–389.e17. b.
18. Madill C, Nguyen DD, McCabe P, et al. Comparison of voice onset measures with glottal pulse identification in acoustic signals: preliminary analyses. *13th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research*. Montréal, Canada; 2019.
19. Kreiman J, Gerratt BR, Kempster GB, et al. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hearing Res*. 1993;36:21–40.
20. Hirano M. *Clinical Examination of Voice*. New York, NJ: Springer; 1981.
21. Yamaguchi H, Shrivastav R, Andrews ML, et al. A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatr Logop*. 2003;55:147–157.
22. Eadie TL, Baylor CR. The effect of perceptual training of inexperienced listeners' judgements of dysphonic voice. *J Voice*. 2006;20:527–544.
23. Roy N, Barkmeier-Kraemer J, Eadie T, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol*. 2013;22:212–226.
24. Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx*. 2015;42:183–188.
25. Peters FMH, Boves L, Van Dielen ICH. Perceptual judgement of abruptness of voice onset in vowels as a function of the amplitude envelope. *J Speech Hear Disord*. 1986;51:299–308.
26. Van de Weijer JC, Slis IH. Nasaliteitsmeting met de nasometer [Nasality measurement with the nasometer]. *Tijdschrift Logopedie Foniatrie*. 1991;63:97–101.
27. Kacha A, Grenez F, Schoentgen J. Voice quality assessment by means of comparative judgements of speech tokens. *9th Interspeech Conference*. Lisbon, Portugal; 2005.
28. Yiu EM, Chan KM, Mok RS. Reliability and confidence in using a paired comparison paradigm in perceptual voice quality paradigm. *Clin Linguist Phon*. 2007;21:129–145.
29. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am*. 2008;122:2354–2364.
30. Maryn Y, Dick C, Vandenbruaene C, et al. Spectral, cepstral, and multivariate exploration of tracheoesophageal voice quality in continuous speech and sustained vowels. *Laryngoscope*. 2009;119:2384–2394.
31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
32. Portney LG, Watkins MP. *Foundations of Clinical Research: Application to Practice*. Upper Saddle River, NJ: Prentice-Hall; 2000.
33. Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Lang Hear Res*. 2011;54:430–447.
34. Rees M. Harshness and glottal attack. *J Speech Hear Res*. 1956;1:344–349.
35. Revis J, Giovanni A, Triglia JM. Influence of voice onset on the perceptual analysis of dysphonia. *Folia Phoniatr Logop*. 2002;54:19–25.
36. Heuer R, Towne C, Hockstein NE, et al. The Towne-Heuer reading passage – a reliable aid to the evaluation of voice. *J Voice*. 2000;14:236–239.