Phonation stabilisation time as an indicator of voice disorder

Felix Schaeffler, Janet Beck & Stephen Jannetts
CASL Research Centre, Queen Margaret University Edinburgh, Scotland, UK
fschaeffler@qmu.ac.uk; sjannetts@qmu.ac.uk; jbeck@qmu.ac.uk

ABSTRACT

There is increasing emphasis on use of connected speech for acoustic analysis of voice disorder, but the differential impact of disorder on initiation, maintenance and termination of phonation has received little attention. This study introduces a new measure of dynamic changes at onset of phonation during connected speech, phonation stabilisation time (PST), and compares this measure with conventional analysis of sustained vowels.

Voice samples obtained from the KayPENTAX Disordered Voice Database were analysed (202 females, 128 males) including 'below threshold' voices where there was a clinical diagnosis but acoustic parameters for sustained vowels were within the normal range.

Female disordered voices showed significantly longer PST duration than normal voices, including those in the 'below threshold' group. Overall differences for male voices were also significant. Results suggest that, at least for females, PST measurement from connected speech could provide a more sensitive indicator of disorder than traditional analysis of sustained vowels.

Keywords: clinical phonetics, voice disorder, acoustic analysis, connected speech

1. INTRODUCTION

The production of connected speech is a highly coordinated and complex process requiring, amongst other things, initiation, maintenance and termination of phonation in precise alignment with rapid articulatory movement. Voice disorders may affect any or all of these aspects of phonation.

Impaired maintenance of phonation, i.e. the ability to sustain adequately periodic vocal fold vibrations [13], is a common feature of voice disorder and has been the focus of many acoustic assessment procedures [23]. These typically analyse short term deviations from periodicity in vocal fold vibration (e.g. shimmer and jitter) over a certain amount of time. This type of analysis is usually based on sustained vowels in order to exclude confounding factors such as consonantal context and articulatory movement [30]. Initial and final portions of the sustained vowel are usually excluded in order to further minimise confounding factors, so that acoustic analysis focuses on the relatively stable mid-portion of a sustained vowel [29].

This approach has been criticised for poor validity and for exclusion of factors that may be highly relevant for assessing voice disorder [1, 6, 14, 24, 25, 28]. The complex transitions required in connected speech could be a rich source of clinically relevant data, because the mechanical consequences of minor inflammatory changes or muscle tension may be most evident at voice onset, when the destabilising effects of increased inertia or stiffness will be greatest [16, 31]. Unfortunately, these transitions are explicitly excluded by typical protocols for acoustic analysis of sustained vowels and, even where acoustic analysis does involve connected speech, the initiation, maintenance and termination of phonation are very rarely differentiated.

A small number of studies have used laryngographic techniques to look at concepts like vocal attack time [2, 12, 26] or other details of voice initiation. In this context, Fourcin and Abberton [10] observed substantial disturbance in the laryngograph signal of disordered voice at voicing onset even where normal periodicity was achieved during sustained vowel production, supporting the notion that voicing initiation might be impaired even if voicing maintenance is within the normal range.

Purely acoustic approaches to clinical analysis of voicing onset are rare [21] and those that exist, such as vocal rise time (VRT, see [3], p. 129 and references there), have not found widespread use, maybe due to signal processing challenges at the time of their development. However, modern signal processing hardware and software allow rapid analysis of large amounts of data, and there is now a strong case to support routine inclusion of (a) connected speech and (b) initiation and termination of vocal fold vibrations in clinical voice assessment.

Many perceptual tools for clinical voice analysis (e.g GRBAS [17], Vocal Profile Analysis [22] and CAPE-V [20]) are designed to evaluate connected speech and thus, at least implicitly, take initiation and termination into account. Most acoustic approaches, however, still focus mainly on sustained vowels. Maryn et al. [23] conducted a meta-analysis of studies evaluating acoustic measures and their correlations with perceptual measures of overall voice quality. Of those studies which met inclusion criteria, 21 used sustained vowels compared with

seven using continuous speech samples. This meta-analysis suggested that alternative acoustic parameters (e.g. Qi et al.'s signal-to-noise ratio [27] and Hillenbrand et al.'s cepstral peak prominence [15] measures) lend themselves much better to the analysis of connected speech than traditional parameters like shimmer and jitter. This may be because these measures do not rely on accurate determination of the glottal cycle.

In the present study we focussed on the analysis of periodicity patterns at the onset of voiced portions of connected speech signals. We analysed the behaviour of the autocorrelation values over a 12s connected speech extract, in order to measure the time taken for the autocorrelation values to rise from a voicing threshold to a "stable periodicity threshold". The voicing threshold was set at .45, following the standard of our chosen acoustic analysis software Praat [4], and the "stable periodicity threshold" was set at .91, following pilot testing (see below). The time between voicing threshold and "stable periodicity threshold" was called "phonation stabilisation time" (PST). This time was measured for all voiced portions in the signal that were continuously above the voicing threshold for 70 ms or more.

We had three main hypotheses: (a) that disordered voices would show longer mean PST than normal voices; (b) that the standard deviation of PST would be higher in disordered voices; and (c) that the percentage of voiced portions of 70 ms or more that do not reach the stable periodicity threshold would be higher in disordered than normal voices. Note that the last measure does not focus on phonation initiation, but measures a loss in overall periodicity, comparable to cepstral analyses of voice [11].

The main purpose of investigating phonation stabilisation in connected speech was to show whether this approach could reveal deviant patterns in disordered voices that are not picked up by conventional acoustic analysis of sustained vowels. For this reason we analysed not only voices which were clearly identifiable as "normal" and "disordered", but also a subset of voices where there was a clinical diagnosis but acoustic analysis of sustained vowels showed that all acoustic parameters were in the normal range.

2. METHOD

2.1. Material

The voice samples for this study were taken from the KayPENTAX Disordered Voice Database [8]. This database contains data from about 700 speakers (classified as normal or by diagnostic category). For most speakers, a sustained vowel and connected speech sample are provided, together with information about native language, gender, age, smoking status and diagnostic notes.

For the current study we selected samples from all speakers where: (a) audio data included a connected speech sample as well as a sustained vowel, (b) English was the native language, and (c) the sample had a valid diagnostic label. This resulted in a selection of 330 samples; see Table 1 for gender and health state distribution.

Table 1: Number of analysed samples by gender and health state

	Normal	Disordered	Total
Female	32	170	202
Male	21	107	128
Total	53	277	330

The connected speech samples in the database consist of the first 12 seconds of the 'Rainbow passage' [9]. The amount of phonetic material contained in the sample depends on the speech rate of the speaker. The samples were not edited before processing.

2.2. Acoustic analysis

Acoustic analysis was performed with Praat 5.4.04. The sound files contained in the Disordered Voice Database differ in sampling frequency. Some were sampled with 25 kHz, others with 10 kHz. As the sampling frequency has potential influence on the analytical procedure outlined below, all 25 kHz files were re-sampled to 10 kHz before analysis.

Each sound file was initially analysed with Praat's 'To Pitch (ac) …' function [5], using Praat's standard settings for this function, apart from 'pitch ceiling', which was set to 500 Hz. From the resulting 'Pitch object', a segmentation into voiced and unvoiced parts of the signal was derived, using two of Praat's standard functions ('To PointProcess (cc)' and 'To TextGrid(vuv)', both with standard settings - see Praat manual [4] for details).

Autocorrelation values were derived from the 'Pitch object' by extracting the highest autocorrelation value within each pitch frame for frequencies between 75 and 500 Hz (note that Praat does not give direct access to autocorrelation values, these were derived from a 'text file' version of the 'Pitch object').

Voiced segments of 70 ms or longer were considered for further analysis. The limit of 70 ms was derived from reports of mean durations of short vowels and nasals in corpora of connected speech [7].

For these sections, the time distance between the beginning of the voiced segment and the threshold

value of 0.91 was determined. This duration will be called 'phonation stabilisation time' (PST) in the following text.

The threshold value of 0.91was determined in a pilot study. 10 normal speakers were randomly selected from the database, voiced portions were manually selected and the maximum autocorrelation value for each voiced portion was determined. The value of 0.91 constituted the approximate lower quartile of the distribution and was thus chosen as a value that was reached by most typical speakers in the sample most of the time.

Not all voiced segments of 70 ms or longer reached the stable periodicity threshold at some point during the segment. These segments were not included in PST calculations. The proportion of voiced segments reaching threshold for every speaker was recorded as a percentage (cf. hypothesis (c) above). This variable will be called 'Seg%' in the following.

2.3. Selection of 'below threshold voices'

The Disordered Voice Database includes values for various acoustic parameters, measured for the sustained vowel samples. These parameters were derived with the Multidimensional Voice Program (MDVP) [19], and the MDVP handbook provides normative thresholds for 28 parameters, 22 of which are reported for the sustained vowel samples in the Disordered Voice Database.

To identify voices without pathological findings we initially excluded all samples that were above any of the 22 published thresholds, but this led to the exclusion of a substantial number of normal voices. We therefore analysed confusion matrices for all parameters and excluded parameters with a false positive rate above 20% of samples. This led to the exclusion of three MDVP parameters (VAM, VFO and VTI). We also excluded the parameters FTRI and ATRI, because many samples had missing values for these parameters.

26 of the 32 normal female voices were below all remaining thresholds, four violated one or two thresholds and the remaining two violated several thresholds. 11 of the 21 normal male voices were below all thresholds, four violated one or two thresholds and the remaining six violated several thresholds. We therefore assumed that the criterion 'below all thresholds' was still too strict, and included all speakers that violated fewer than three thresholds in the 'below threshold' group.

2.4. Statistical Analysis

In order to test the main hypotheses, the two gender groups were analysed separately. The three dependent variables were compared with t-tests or Mann-Whitney U-tests for the independent variable of health state. Nonparametric tests were chosen when Shapiro-Wilk tests suggested deviations from normality in a sample.

Arcsine transformation was applied to proportion data before applying statistical tests. Non-transformed descriptive measures are provided, if not indicated otherwise.

3. RESULTS

3.1. Female group

One female voice was excluded from further analysis as the autocorrelation threshold was never reached in the sample. Female disordered voices showed a significantly longer average PST than female normal voices [U=968, p<.001]. PST SD was also significantly larger in the disordered group [U=1097, p<.001].

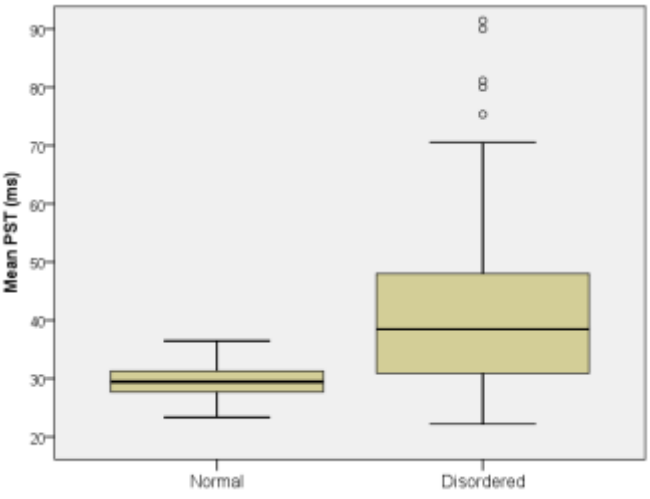
There also was a significantly higher proportion of voiced segments reaching threshold in the normal group than in the disordered group [U=1494, p<.001]. Descriptive values for all three measures are given in Table 2.

Table 2: Female group means (SD) for the three variables under study.

	Normal	Disordered
PST M	29.5ms (2.8)	42.8 ms (18.8)
PST SD	13.0ms (4.2)	26.2 ms (17.7)
Seg %	98.7% (2.0)	89.8% (15.3)

The boxplot in Figure 1 illustrates the distribution for PST mean in both female groups.

Figure 1: Boxplots of mean PST for female normal and disordered groups. Two extreme outliers (disordered, 125ms, 188ms) not shown here.



The trend for higher values in the disordered group is clearly illustrated in Figure 1, but it can also be seen that there is considerable overlap, and the variation in the disordered group is much larger.

3.2. Female group below thresholds

30 normal female voices and 20 disordered female voices fell in the 'below thresholds' group, as defined above. Even for this sub-selection of voices, there was a significant difference in PST mean duration [t (23.19) = -2.619, p<.05, equal variances not assumed] and in PST SD [U=193, p<.05]. Differences in Seg% were not significant. Descriptive values for all measurements are provided in Table 3.

Table 3: Female 'below threshold' group mean (SD) for the three variables under study.

	Normal	Disordered
PST M	44.5 (12.7)	56.1 (19.7)
PST SD	41.5 (10.7)	42.4 (17.5)
Seg %	89.6 (5.8)	86.2 (8.9)

3.3. Male group

One male voice was excluded from further analysis as the autocorrelation threshold of .91 was never reached in the analysed voiced segments. There was a significant effect of Seg% in the male group [U=758, p<.05]. The difference in mean PST was close to significant [U=818.5, p=.056]. PST *SD* differences were not significant. Descriptive values for the male group are given in Table 4.

Table 4: Male group mean (SD) for the three variables under study.

	Normal	Disordered
PST M	40.8 (8.7)	51.1 (20.9)
PST SD	26.1 (11.6)	32.3 (22.4)
Seg %	94.3 (6.2)	85.2 (16.9)

These values indicate that male disordered voices show similar effects to female voices but the outcomes are less clear, apart from the difference in Seg%. The male 'below threshold' classification resulted in a sub-selection of 15 normal and 17 disordered voices. None of the differences between these two groups reached significance.

4. DISCUSSION

This paper presents a new method for clinical voice analysis that focusses on dynamic changes in phonation during connected speech, rather than on periodicity in sustained vowels. The results indicate that PST could add important information about vocal fold function. PST differences between normal disordered voices were much clearer for females, and the results for the female 'below threshold' group were particularly striking; they suggested that PST might be an indicator of pathology even in cases of voice disorder where acoustic parameters derived from sustained vowels are within the normal range. PST appears to be a more sensitive measure of pathology, meaning that it could have potential in early recognition of voice problems. This is in line with our expectation that the vibratory consequences of developing changes within the larynx will be evident first at voicing transitions.

This study applied PST measurement to pre-existing acoustic material without any major pre-processing. While this led to significant results, which is promising in terms of robustness and economy, the approach will require further scrutiny before any clinical application can be attempted.

For example, future studies should evaluate the effects of segmental context on PST as there is evidence that preceding segments can influence F0 height and perturbation at the onset of voiced segments [21]. These effects should not only be considered as potentially confounding factors but might also interact with voice pathology. Clarifying these effects and interactions would aid the design of test material (reading passages or similar) with optimum diagnostic value.

Details of the signal processing procedure used here should also be analysed further. So far we have mainly relied on standard settings of Praat [4], and fine-tuning of parameters might improve the specificity and sensitivity of the tool. The chosen periodicity criterion of .91 will also require review. This seems to be appropriate for female voices, but its behaviour with male voices is less convincing. A lower criterion for male voices might lead to better differentiation.

The division into 'normal' and 'disordered' voices is also somewhat over-simplistic. Preliminary analysis of diagnostic information from the Disordered Voice Database shows some patterning, but diagnostic information in the database is not always entirely coherent. For example, some voices have been given up to six different diagnostic labels. The database also lacks detail about severity of disorder, onset of disorder or client evaluations of impact. Future development and evaluation of PST would benefit from a more finely tuned differentiation of disorder type and severity, including careful laryngoscopic descriptions. If future studies can confirm the potential of PST, we will also need extensive normative data, including information about typical within-speaker variation.

5. ACKNOWLEDGMENTS

We would like to thank our student Sarah Bruce for valuable help with the pilot study.

6. REFERENCES

[1] Askenfelt, A.G. & Hammarberg, B. 1986. Speech waveform perturbation analysis: a perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research*, 29(1), 50–64.

[2] Baken, R. & Orlikoff, R. 1998. Estimating vocal fold adduction time from EGG and acoustic records. In Programme and abstract book: *24th IALP Congress*, Amsterdam.

[3] Baken, R. J., & Orlikoff, R. F. 2000. *Clinical Measurement of Speech and Voice.* San Diego: Singular Publishing Group.

[4] Boersma, P. & Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 28 December from http://www.praat.org.

[5] Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences* 17. University of Amsterdam, 97-110.

[6] Choi, S.H. et al. 2012. The effect of segment selection on acoustic analysis. *Journal of Voice*, 26(1), 1–7.

[7] Crystal, T. H., & House, A. S. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, 83(4), 1553-1573.

[8] Disordered Voice Database, model 4337. 1994. developed by the Massachusetts Eye and Ear Infirmary Voice and Speech Lab. *KayPENTAX Corporation*, Montvale, NJ, Version 1.03.

[9] Fairbanks, G. 1960. *Voice and Articulation Drill Book* (2nd ed.). New York: Harper.

[10] Fourcin, A. & Abberton, E. 2008. Hearing and phonetic criteria in voice measurement: clinical applications. *Logopedics, Phoniatrics, Vocology*, 33(1), 35–48.

[11] Fraile, R. & Godino-Llorente, J.I. 2014. Cepstral peak prominence: A comprehensive analysis. *Biomedical Signal Processing and Control*, 14, 42–54.

[12] Francis, A.L., Ciocca, V. & Yu, J.M.C. 2003. Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America*, 113(2), 1025–32.

[13] Gordon, M. & Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4), 383–406.

[14] Hammarberg, B. et al. 1980. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-Laryngologica*, 90(5-6), 441–51.

[15] Hillenbrand, J. & Houde, R.A. 1996. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing research*, 39(2), 311–21.

[16] Hirano, M. & Bless, D.M, 1993. *Videostroboscopic examination of the larynx.* San Diego: Singular Publishing Group.

[17] Hirano, M. 1981. Psycho-acoustic evaluation of voice: GRBAS Scale for evaluating the hoarse voice. In: *Clinical Examination of Voice*. London: Springer, 81–94.

[18] Hombert, J. M., Ohala, J. J., & Ewan, W. G. 1979. Phonetic explanations for the development of tones. *Language*, 37-58.

[19] KayPENTAX. 2008. Operations Manual: Multi-dimensional Voice Program (MDVP) Model 5105. Lincoln Park, NJ: KayPENTAX.

[20] Kempster, G.B. et al., 2009. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18(2), 124–32.

[21] Koike Y., 1973. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica* VII, 17-23.

[22] Laver, J., Wirz, S., Mackenzie, J. & Hiller, S.M.,1991. A perceptual protocol for the analysis of vocal profiles. In J. Laver, (Ed.) *The Gift of Speech*. Edinburgh: Edinburgh University Press, 265-280.

[23] Maryn, Y. et al., 2009. Acoustic measurement of overall voice quality: A meta-analysis. *Journal of the Acoustical Society of America*, 126(5), 2619–2634.

[24] Maryn, Y. & Roy, N., 2012. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. *Jornal da Sociedade Brasileira de Fonoaudiologia*, 24(2), 107–12.

[25] Maryn, Y. et al., 2010. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *Journal of Voice*, 24(5), 540–55.

[26] Orlikoff, R.F. et al., 2009. Validation of a glottographic measure of vocal attack. *Journal of Voice*, 23(2), 164–8.

[27] Qi, Y., Hillman, R.E. & Milstein, C., 1999. The estimation of signal-to-noise ratio in continuous speech for disordered voices. *The Journal of the Acoustical Society of America*, 105(4), 2532–5.

[28] Takahashi, H. & Koike, Y., 1976. Some perceptual dimensions and acoustical correlates of pathologic voices. *Acta Oto-Laryngologica. Supplementum*, 338, 1–24.

[29] Titze, I.R., 1995. *Workshop on acoustic voice analysis: Summary statement*. Denver: National Center for Voice and Speech.

[30] Wolfe, V., Cornell, R. & Fitch, J., 1995. Sentence/vowel correlation in the evaluation of dysphonia. *Journal of Voice*, 9(3), 297–303.

[31] Zhang, Z., Neubauer, J. & Berry, D.A., 2007. Physical mechanisms of phonation onset: a linear stability analysis of an aeroelastic continuum model of phonation. *The Journal of the Acoustical Society of America*, 122(4), 2279–95.