VOICE QUALITY IN NORWEGIAN: EFFECTS OF DIALECT, SPEAKER AGE, SEX, AND SPEAKING MODE

Wim A. van Dommelen

Department of Language and Literature, NTNU, Trondheim, Norway
wim.van.dommelen@ntnu.no

ABSTRACT

This study investigates the effect of Norwegian speakers' dialectal background, age, and sex as well as speaking mode on voice quality. Recordings of read and semi-spontaneous speech were used to analyze short and long /a/ vowels produced by younger (< 40 years) and older (> 40 years) female and male speakers from four main dialectal variants. Acoustic analysis comprised both spectral tilt and harmonicity measures. The analysis revealed that the factor dialect did not have a robust impact on any voice quality measure. Speaker sex affected both types of measures more than any of the other factors, females having steeper spectral tilt and higher harmonicity values than men. Only the spectral measures differed with age, older speakers demonstrating steeper slopes than younger subjects. For spontaneous speech, somewhat steeper slopes and higher harmonicity values were observed than for read speech. Current results conflicting with previous evidence are discussed.

Keywords: Voice quality, Norwegian dialects, speaker age, sex, speaking mode

1. INTRODUCTION

The goal of this study is to explore the relative contribution of Norwegian speakers' dialectal background, age, and sex as well as speaking mode to voice quality. Previous research on dialect-specific voice quality is scarce. Additionally, the presented evidence is not always conclusive. Investigating breathiness in normal female speech, Henton & Bladon [13] measured H1 and H2 in speakers of Received Pronunciation and Modified Northern English. The effect of dialectal accent was not evaluated, but it was small and probably not robust. Szakay [28] investigated phonation in spontaneous speech produced by speakers of Maori English and Pakeha English. H1-H2 values for the former group were found to be generally lower, but no values are specified. Esling & Wong [6] characterize Norwegian voice quality as whispery creaky, but they do not mention any dialectal differences.

The picture emerging from research on the effect of age on phonation is not consistent. While Lee et al. [18] observed H1-H2 and H1-A1 to be smaller for elder than for younger speakers, no significant effect for H1-A1 was found by Gorham-Rowan & Laures-Gore [10]. The latter reported higher noise levels for older speakers (similarly Jin et al. [14], Lortie et al. [19]; see also the meta-analysis by Rojas, Kefalianos & Vogel [24]). In contrast, no age-related changes in noise levels were found by Awan [1] and Goy et al. [11].

Likewise, reported results regarding the impact of speaker sex are confusing. Across publications, all measures apart from H1-H2 (indicating steeper slopes for females) and harmonicity (indicating less noise for females) produced inconclusive evidence. Larger H2-H4 values for females were found by Garellek et al. [9], the opposite by Hejná et al. [12]. The latter measured no difference for H1*-A1*, but lower female CPP compared to higher female CPP values found by Garellek & Keating [8]. Data presented by Mezzedimi et al. [21] and Schaeffer, Knudsen & Small [25] seem to show no impact of speaker sex on HNR / NHR, but no statistical tests were performed.

The two speaking modes investigated in the current study are read speech and semi-spontaneous speech. Studies including the influence of speaking mode on phonation seem to be extremely rare. Participants in Lortie et al. [19] produced a sustained vowel /a/ and retold popular story tales. Acoustic analysis revealed HNR to be lower in connected speech. It is important to note that the latter type of speech contained both vowels and voiced consonants. Starr [26] analyzed Japanese so-called sweet and non-sweet voice performances revealing higher values for sweet voice for H1-H2, H1-A1, and H1-A3. The results for HNR were not consistent across speakers, and no consistent difference was found for CPP.

Due to the scarce and partially contradictory evidence on factors influencing voice quality, the present study has a heuristic character. The two main research questions are:

- What is the effect of speakers' dialectal background, age, sex, and speaking mode (read vs. semi-spontaneous speech) on spectral tilt and harmonicity measures?
- What is the effect size of each of these factors?

2. METHOD

2.1. Speech material

Materials used for this investigation were chosen from the speech database NB Tale [22] provided by the National Library of Norway. The corpus contains annotated recordings of 240 native speakers of Norwegian. This material is divided according to the speakers' dialectal background into 12 subgroups containing 20 speakers each. For the present purposes four homogeneous groups of 20 speakers each were selected, representing the four main dialects: Northern Norwegian (henceforth: North), Central Norwegian (Central), Western Norwegian (West), and Eastern Norwegian (East). In addition to dialect, speaker age (18-40 yrs vs. 41-80 yrs) and sex (female vs. male) are included as systematic factors in the database. Within each subgroup, there are thus five younger and five older speakers of each sex. All recordings were made in a sound-treated environment using 48 kHz sampling frequency and 16-bit quantization. Recordings were in stereo using a Sennheiser 2-1-5 headset and a Shure KSM 44 studio microphone. Since the frequency characteristic of the latter is dependent on the distance from the speaker to the microphone, for the present acoustic analysis the Sennheiser headset channel was selected.

The NB Tale database contains read sentences and semi-spontaneous speech. The read material consists of 20 sentences per speaker. While the first three sentences are identical for all speakers, the remaining 17 have different content across speakers. From the sentence material, stressed /a(:)/ vowels were selected, five from sentences 1-3 that were the same for all speakers and varying numbers of /a(:)/ vowels that were contained in sentences 4-20. The spontaneous speech material contains recordings of participants speaking approximately two minutes about a topic of their own choice. From this material, /a(:)/ vowels and (to increase the number of available tokens) some /æ(:)/ vowels were selected. Our aim was to analyze a total number of 20 spontaneously produced vowel tokens per speaker, but for three out of 80 subjects fewer than 20 tokens were available.

2.2 Acoustic analysis and further processing

Using a Praat [3] script, vowel tokens were extracted from the speech recordings. The freely available tool PraatSauce [15] was used for spectral analysis. Five spectral tilt parameters and five harmonicity measures were calculated: H1*-H2*, H2*-H4, H1*-A1*, H1*-A2*, and H1*-A3*; and CPP (Cepstral Peak Prominence), HNR05, HNR15, HNR25, and HNR35; all in dB. Asterisks denote that PraatSauce outputs for spectral magnitudes of H_i and A_i were corrected for the effect of formant frequencies and bandwidths. (The tool does not correct H4.) To minimize possible effects of phonetic context on measurements, the central 20 ms portion of each vowel was selected for subsequent evaluation. Since an analysis frame rate of 1 ms was used, this corresponds to 20 frames per vowel token for all extracted measures. To eliminate analysis errors, further selection was performed in two steps.

a) All cases with erroneous f0 values of 0 Hz were excluded from the raw data.
b) Subsequently, formant values for all /a(:)/ vowels had to comply with mean values +/- 2 sd as measured for females and males in NB Tale.

Consequently, for a number of vowel tokens there were less than 20 frame values. Remaining frame values were used to calculate mean values for each of the measured parameters. The total number of tokens that went into statistical analysis was 4105 (read speech: 2712; spontaneous speech: 1393).

2.3 Statistical analysis

Statistical analysis of the data was performed with the R program's package lme4 [23] calculating Linear Mixed Effects Models (LMEM). LMEMs included Dialect (North, Central, West, East), Age (< 40, > 40), Sex (female, male) and Mode (read, spontaneous) as fixed factors with by-subject random slopes and intercepts for the factor Mode [2]. Following calculation of each LMEM, the step function developed by Kuznetsova, Brockhoff & Christensen [16] was used to perform backward elimination of nonsignificant factors. Likelihood ratio tests were run in each case comparing a model with n factors with a model having n-1 factors (i.e., without the factor under scrutiny).

Results revealed that the factor Dialect never reached statistical significance, but the other three fixed factors (Age, Sex and Mode) often did. Following statistical analysis, effect size *d* for the levels of the latter three factors was calculated as the absolute difference between two level means divided by the square root of the sum of the variances of the intercept per subject, slope for Mode per subject and the residual variance (Brysbaert & Stevens [4]; Westfall, Kenny & Judd [29]). Calculations of *d* involved both statistically significant and nonsignificant effects.

3. RESULTS

3.1 Main factors

As already stated above, for none of the measures investigated did the factor Dialect reach statistical significance. Between-group differences for the other factors are presented in Figure 1. As to the effect of Age, for speakers over 40 years significantly higher values than for younger speakers were found for three out of five spectral measures (between-group differences: H1*-H2*: 1.9 dB, H1*-A1*: 1.2 dB, and H1*-A2*: 1.4 dB; $p < 0.001$, $p = 0.025$, and $p = 0.019$, respectively). In contrast, the two age groups did not differ with respect to harmonicity measures. More comprehensive effects were observed for the factor Sex, where except for H2*-H4 female subjects had higher values than males. All spectral measures apart from H1*-A1* as well as all harmonicity measures reached statistical significance (for H2*-H4, p = 0.007; for H1*-A2*, p = 0.031; else $p < 0.001$). For the spectral measures, significant between-group differences ranged between -0.9 dB for H2*-H4 and 4.9 dB for H1*-A3*; harmonicity measures varied between 3.0 dB for CPP and 10.2 for HNR05. Also speaking mode appeared to affect voice quality in several parameters with values for spontaneous speech generally being higher than for read speech. Three out of five spectral measures were higher for read than spontaneous speech (H2*-H4: 0.3 dB, p = 0.047; H1*-A1*: 0.5 dB, p = 0.003; and H1*-A3*: 0.7 dB, p = 0.004), as well as all four HNR measures (HNR05: 1.1 dB, HNR015: 1.1 dB, HNR25: 1.0 dB, and HNR35: 1.0 dB; in all cases $p < 0.001$).

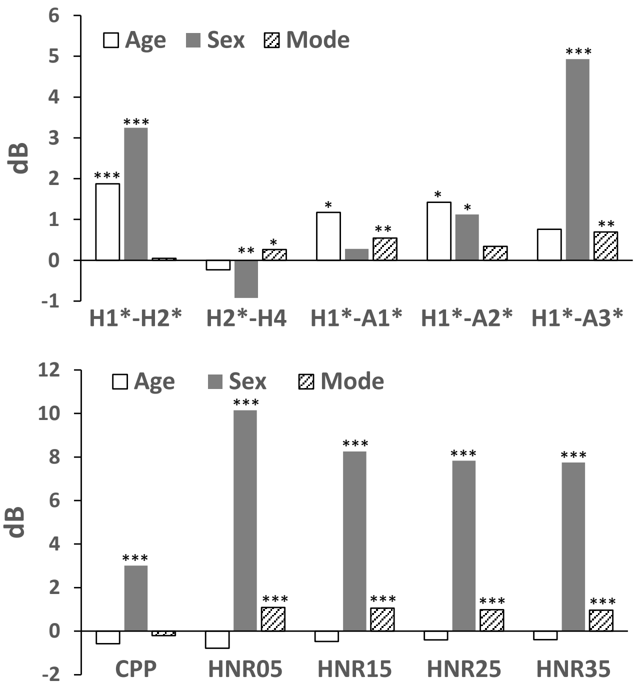

Figure 1: Between-group differences for spectral (top panel) and harmonicity (bottom panel) measures. Age: older (> 40 yrs) - younger (< 40 yrs); Sex: females - males; Mode: spontaneous - read. Probability levels of 0.05, 0.01 and 0.001 are indicated by *, **, and ***, respectively

3.2 Effect sizes

It is evident from Figure 1 that the factor Sex has the largest impact on voice quality, for spectral measures mainly due to H1*-H2* and H1*-A3*. Less important is the factor Age, followed by Mode. To quantify this hierarchy, for each of the acoustic measures effect size *d* was calculated (see Section 2.3). As can be seen from Table 1, across spectral measures, the Sex/Age effect ratio amounted to 0.36/0.21 = 1.8, while the Sex/Mode ratio was even higher (0.36/0.07 = 5.0; rounded values). Among harmonicity measures, the factor Mode had a more consistent impact, and the Sex/Mode *d* ratio was even somewhat lower than the one for Sex/Age (1.25/0.14 = 8.8 vs. 1.25/0.10 = 13.0). Averaged across all acoustic measures, Sex affected voice quality most, followed by Age, with Mode being somewhat less influential (effect ratio for Sex/Age: 0.80/0.15 = 5.3 vs. Sex/Mode: 0.80/0.11 = 7.5).

3.3 Interactions

For reasons of space, this section will only mention some main interaction results. Merely 12 out of 60 two-way interactions turned out to be significant. Five Age x Sex interactions reached significance (for H1*-H2*, and HNR05 – HNR35), all showing larger sex effects for younger vs. older speakers. For four out of five spectral measures, significant Dialect x Sex interactions were observed (H1*-H2*, H1*-A1*, H1*-A2*, and H1*-A3*), in addition to CPP, all five being seemingly incidental. Additionally, one Sex x Mode (H2*-H4), and one Dialect x Mode interaction (HNR05) reached statistical significance.

4. DISCUSSION

The present study investigated the effect of Norwegian speakers' dialectal background, age, and sex as well as speaking mode on voice quality. Acoustic measures characterizing voice quality comprised spectral tilt parameters H1*-H2*, H2*-H4, H1*-A1*, H1*-A2*, H1*-A3* and harmonicity parameters CPP, HNR05, HNR15, HNR25, and HNR35.

4.1 Dialect

Altogether, the picture emerging from the measurements was relatively consistent. For a start, the factor dialect appeared to have no systematic impact on any voice quality measure. This result seems to be in line with the small difference in H1-H2 for two English dialects found by Henton & Bladon [13]. It could be speculated that voice quality

Table 1: Effect size *d* for spectral measures H1*-H2* to H1*-A3* and harmonicity measures CPP to HNR35. Calculations of *d* involved both statistically significant and nonsignificant effects.

Factor	H1*-H2*	H2*-H4	H1*-A1*	H1*-A2*	H1*-A3*	mean	CPP	HNR 05	HNR 15	HNR 25	HNR 35	mean
Age	0.35	0.07	0.25	0.26	0.10	**0.21**	0.15	0.11	0.08	0.07	0.07	**0.10**
Sex	0.60	0.27	0.06	0.21	0.68	**0.36**	0.80	1.42	1.38	1.32	1.31	**1.25**
Mode	0.01	0.08	0.12	0.06	0.10	**0.07**	0.05	0.15	0.18	0.17	0.16	**0.14**

differences may not – or at best rarely – be exploited between dialects, as opposed to phonemically-contrastive phonation types distinguishing registers (DiCanio [5]) or consonants and vowels (e.g., Esposito, Sleeper & Schäfer [7]).

4.2 Speaker age

As to the effect of age on voice quality, the present finding of steeper spectral slopes for older speakers (particularly H1*-H2*) would suggest more breathiness in that age category. This outcome is at odds with the data for Korean female speakers in Lee et al. [18], who found steeper slopes indicated by H1-H2 and H1-A1 for younger women. The authors speculate that the relatively high formant frequencies for Korean /a/ might be partially responsible for their results and that the use of corrected measures could contribute to the interpretation of their results. The current absence of a significant effect of age on harmonicity measures is at variance with the evidence of increased noise in older compared to younger voices in previous investigations ([14], [19], [21], [25]). It seems possible to explain the discrepancies in harmonicity between the present and previous studies by differences in age ranges. The data for SNR presented in Stathopoulos, Huber & Sussman ([27]: Figure 3) show variation but no clear rising or falling tendency for speakers aged 20 – 70 years. For older speakers (70 – 93 years), SNR decreases with age. Whereas in the investigations cited above, age distributions typically range between 20 and 85 years, the current age range was 18 – 71 years, thus excluding ages where decreasing harmonicity could be expected. As we have seen above, our speakers did show an effect of age on spectral slopes. This indicates that spectral slopes and harmonicity measure different acoustic properties of the glottal waveform, breathiness and roughness or hoarseness, respectively (Hejná et al. [12], Latoszek et al. [17], Yumoto, Gould & Baer [30]).

4.3 Speaker sex

With some exceptions, the present effect of speaker sex on acoustic measures parallels previous findings. All previously mentioned studies reported steeper H1*-H2* and H1*-A3* slopes for females. The current higher H2*-H4 values for male speakers are in line with Hejná et al. [12] and Matar et al. [20] but at odds with Garellek et al. [9]. Even less consistent are the results for women's lower H1*-A1* in Gorham-Rowan & Laures-Gore [10] and H1*-H2* and H1*-A1* values in Garellek & Keating [8]. These incongruities may be explained by differences in experimental methodologies and technical conditions. In this connection it is noteworthy that the less robust measures in previous investigations achieved the smallest sex-specific effects in the present study. As to harmonicity (HNR) measures, largely confirming results have been found, women's voices containing less noise than men's.

4.4 Speaking mode

As shown by the current data, also speaking mode had a certain impact on voice quality. In spontaneous speech, three spectral tilt measures (H2*-H4, H1*-A1*, and H1*-A3*) and all HNR measures were larger than in read speech. The effects were relatively small but consistent. The fact that Lortie et al. [19] observed an opposite effect for HNR can be due to their use of portions of spontaneous speech, containing both vowels and consonants. Particularly the presence of obstruents may have resulted in lower HNR values compared to isolated vowels.

5. CONCLUSION

The two research questions formulated in the introduction can be answered as follows:

a) Speaker age and sex as well as speaking mode have been demonstrated to affect both spectral tilt and harmonicity measures. Dialectal background does not exert any systematic effect.
b) In the spectral measures, speaker sex had the greatest impact, followed by speaker age, and mode had by far the weakest effect. As to harmonicity, speaker sex had a more marked effect than the other two factors. While both mode and speaker age were far less important, the latter even failed to reach statistical significance.

6. REFERENCES

[1] Awan, S. N. 2006. The aging female voice: Acoustic and respiratory data. *Clinical Linguistics & Phonetics* 20(2/3), 171-180.

[2] Barr, D.J., Levy, R., Scheepers, C., Tily, H.J. 2013. Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 255-278.

[3] Boersma, P., Weenink, D. 2021. *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.48. http://www.praat.org/.

[4] Brysbaert, M., Stevens, M. 2018. Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition* 1(1): 9, 1-20.

[5] DiCanio, C. T. 2009. The phonetics of register in Takhian Thong Chong. *J. Int. Phon. Ass.* 39 (2), 162-188.

[6] Esling, J. H., Wong, R. F. 1983. Voice quality settings and the teaching of pronunciation. *TESOL Quarterly* 17, 89-95.

[7] Esposito, C. M., Sleeper, M., Schäfer, K. 2021. Examining the relationship between vowel quality and voice quality. *J. Int. Phon. Ass.* 51(3), 361-392.

[8] Garellek, M., Keating, P. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *J. Int. Phon. Ass.* 41(2), 185-205.

[9] Garellek, M., Samlan, R., Gerratt, B. R., Kreiman, J. 2016. Modeling the voice source in terms of spectral slopes. *J. Acoust. Soc. Am.* 139(3), 1404-1410.

[10] Gorham-Rowan, M. M., Laures-Gore, J. 2006. Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders* 39(3), 171-184.

[11] Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., van Lieshout, P. 2013. Normative voice data for younger and older adults. *Journal of Voice* 27, 545-555.

[12] Hejná, M., Šturm, P., Tylečková, L., Bořil, T. 2021. Normophonic breathiness in Czech and Danish: Are females breathier than males? *Journal of Voice* 35(3), 498.e1-498.e22.

[13] Henton, C. G., Bladon, R. A. W. 1985. Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication* 5(3), 221-227.

[14] Jin, S. M., Kwon, K. H., Kang, H. G. 1997. Acoustic and stroboscopic characteristics of normal person's voices with advancing age. *Journal of The Korean Society of Laryngology, Phoniatrics and Logopedics* 8, 44-48.

[15] Kirby, J. P. 2021. [Spectral analysis tool] https://github.com/kirbyj/praatsauce.

[16] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13), 1-26.

[17] Latoszek, B. B. v., Maryn, Y., Gerrits, E., De Bodt, M. S. 2018. A Meta-Analysis: Acoustic Measurement of Roughness and Breathiness. *J. Speech Lang. Hear. Res.* 61(2), 298-323.

[18] Lee, S. J., Cho, Y., Song, J. Y., Lee, D., Kim, Y., Kim, H. 2015. Aging Effect on Korean Female Voice: Acoustic and Perceptual Examinations of Breathiness. *Folia phoniatrica et logopaedica* 67(6), 300-307.

[19] Lortie, C. L., Thibeault, M., Guitton, M. J., Tremblay, P. 2015. Effects of age on the amplitude, frequency and perceived quality of voice. *Age* 37(6), 117.

[20] Matar, N., Portes, C., Lancia, L., Legou, T., Baider, F. 2016. Voice quality and gender stereotypes: A study of Lebanese women with Reinke's edema. *Journal of speech, language, and hearing research* 59(6), S1608-S1617.

[21] Mezzedimi, C., Di Francesco, M., Livi, W., Spinosi, M. C., De Felice, C. 2017. Objective evaluation of presbyphonia: Spectroacoustic study on 142 patients with Praat. *Journal of Voice* 31(2), 257.e25-257.e32.

[22] NB Tale – en grunnleggende akustisk fonetisk taledatabase for norsk [NB Speech – a fundamental acoustic-phonetic speech database for Norwegian]. http://www.nb.no/sprakbanken/show?serial=sbr-31&lang=nb

[23] R Core Team. 2021. R: A language and environment for statistical computing. Retrieved from https://ww.R-project.org/.

[24] Rojas, S., Kefalianos, E., Vogel, A. 2020. How does our voice change as we age? A systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age. *Journal of Speech, Language, and Hearing Research* 63(2), 533-551.

[25] Schaeffer, N., Knudsen, M., Small, A. 2015. Multidimensional voice data on participants with perceptually normal voices from ages 60 to 80: A preliminary acoustic reference for the elderly population. *Journal of Voice* 29(5), 631-637.

[26] Starr, R. L. 2015. Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society* 44, 1-34.

[27] Stathopoulos, E. T., Huber, J. E., Sussman, J. E. 2011. Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *J. Speech Lang. Hear. Res.* 54(4), 1011-1021.

[28] Szakay, A. 2012. Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. *Journal of Sociolinguistics* 16(3), 382-397.

[29] Westfall, J., Kenny, D. A., Judd, C. M. 2014. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psych.: General* 143(5), 2020-2045.

[30] Yumoto, E., Gould, W. J., Baer, T. 1982. The harmonics-to-noise ratio as an index of degree of hoarseness. *J. Acoust. Soc. Am.* 71, 1544–1550.