Effects of fundamental frequency and harmonics-to-noise ratio on the perception of Danish laryngealized phonation

Jailyn M. Peña

New York University
jailynpena@nyu.edu

ABSTRACT

This study examines the effect of fundamental frequency (F0) and harmonics-to-noise ratio (HNR) on the perception of laryngealized phonation in Danish. Previous research shows that compared to modal phonation, laryngealized phonation in Danish generally begins with high F0 at the beginning of the syllable rhyme, followed by low F0 and creaky phonation at the end of the syllable. In this study, F0 and HNR were manipulated in naturally produced nonce words to examine how high F0, low F0, and HNR contribute to the percept of laryngealization. Results from an ABX task and a perceptual rhyming task show that listeners were more likely to identify words with both (1) high F0 and (2) a drop in pitch or lower HNR on the coda sonorant as 'laryngealized' compared to other words. Words with only one laryngealization cue did not promote 'laryngealized' responses, indicating that listeners utilize multiple acoustic cues to perceive laryngealization.

Keywords: laryngealization, Danish, phonation, voice quality, speech perception.

1. INTRODUCTION

This study investigates the effects of fundamental frequency (F0) and harmonics-to-noise ratio (HNR) on the perception of the Danish phonation contrast. Danish phonologically contrasts modal phonation and a type of non-modal phonation similar to creaky phonation, e.g., *ven* [vɛn] 'friend' vs. *vend* [vɛnˀ] 'turn!' [1–3]. This non-modal phonation (here, [ˀ]) can only occur in stressed syllables with either a long vowel or a vowel followed by a sonorant coda consonant in the standard dialect and some dialects spoken in Zealand. Prior literature has referred to this phonation type as 'stød,' though here it will be referred to as laryngealization, following descriptions in Grønnum et al. [4]. This is to acknowledge non-modal phonation in Danish is not identical to creak but is on the creakier end of the phonation spectrum put forth by Gordon & Ladefoged [5].

Acoustically, differences in phonation are usually quantified using measures of spectral tilt and noise. Laryngealized segments are often produced with a smaller difference between the amplitudes of the first and second harmonics (lower H1*-H2*, where the asterisks indicate correction for formants) and lower HNR, indicating more noise in the acoustic signal compared to modally phonated segments [6–11]. Additionally, laryngealized segments may also have lower F0 than modal ones [12].

A unique feature of Danish is that phonologically laryngealized syllables are acoustically characterized by high F0 and modal phonation at the beginning of the rhyme. This first phase is then followed by a second phase with a sharp drop in F0 and intensity, irregular amplitude, and oftentimes creaky or compressed phonation (that is, a period of truly non-modal phonation). In terms of spectral tilt and noise, previous work has found a connection between H1-H2, HNR, and laryngealization in Danish [13]. However, previous research has shown that the realization of the second phase is highly variable, and no single acoustic cue is present in all instantiations [2, 14]. Furthermore, studies have yet to investigate how the various acoustic cues that differentiate modal and laryngealized phonation in production affect native Danish listeners' perception of this phonological phonation contrast.

In this study, we focus on the effects of F0 and noise, quantified in terms of HNR, on native Danish listeners' perception of the phonation contrast. The effect of spectral tilt is not examined here, as techniques for manipulating it are not as well developed as those for F0 and noise.

2. METHODS

38 native Danish listeners between 18-40 ($\mu = 27.3$, $\sigma = 3.9$) participated in the study in Copenhagen, Denmark. Due to the highly variable nature of laryngealized phonation in Danish, two tasks were conducted to probe different aspects of perception and ensure the reliability of the results. An ABX task was conducted to facilitate a direct comparison of our test stimuli with both modally phonated and laryngealized words. A rhyming task was also conducted as a test of listeners' metalinguistic awareness of laryngealized phonation and their ability to generalize this knowledge to novel words.

All listeners were from the dialectal region of Zealand, Denmark where the distribution of laryngealized syllables in the lexicon is the same as in Standard Danish. Listeners performed the tasks in a quiet room on an Acer Spin 5 laptop wearing Sennheiser HD 400S headphones and recorded their answers using a DirectIN high speed button box. All listeners performed the ABX task before the rhyming task. Three listeners who incorrectly performed the rhyming task were excluded from that analysis; all listeners' data were included in the ABX analysis.

To create the stimuli for the two tasks, eight pairs of monosyllabic C(C)VC nonce words produced with both modal and laryngealized phonation as well as four pairs of real monosyllabic words phonologically contrasting in phonation were recorded in a short nonsense phrase consisting of the target word followed by the word *videre* [viðʌ] 'continue.' The recordings were made by a female native Danish speaker (age 27) using a Tascam DR 40x recorder and a head-mounted Shure SM35 condenser microphone at a sampling rate of 44,100 Hz. Examples are given in Table 1. The real words and the laryngealized nonce words were recorded once, and the modally phonated nonce words were recoded twice for a total of 32 recordings. Examples are given in Table 1.

	Nonce Words	Real Words
[an]	[gan], [ganˀ]	*man* [man] 'but' *mand* [manˀ] 'man'
[al]	[nal], [nalˀ]	*tal* [tʰal] 'number' *bal* [balˀ] 'ball'
[ɛn]	[jɛn], [jɛnˀ]	*ven* [vɛn] 'friend' *vend* [vɛnˀ] 'turn!'
[ɛl]	[nɛl], [nɛlˀ]	*vel* [vɛl] 'well' *væld* [vɛlˀ] 'abundance'

Table 1: Examples of stimuli.

The target words were excised from the following word, and vowel and coda sonorant durations were equalized in Praat [15] to approximately 111 ms each for a total rhyme duration close to 222 ms per recording (1.5 times the original average). This was done because preliminary inspection revealed that the original average rhyme duration sounded unnaturally short in isolation. Minor variations in duration occurred based on the period of the soundwave of the original recordings to prevent acoustic distortions. Onset durations varied from 15.6 – 96.6 ms based on the voicing, manner, and number of consonants. No further manipulations were made to the real words, the laryngealized nonce words, and one set of modally phonated nonce words.

To examine how F0 and HNR affect listeners' perception of the phonation contrast, the second set of modally phonated nonce words (16 recordings) were acoustically manipulated to show one of 6 F0 contours (HH, HL, HM, LL, ML, MM, where 'H' indicates high pitch, 'M' indicates mid pitch, and 'L' indicates low pitch) using a Praat script [16]. The first letter in the sequence indicates the relative F0 of the vowel and the second letter indicates the relative F0 of the coda sonorant. The high F0 value (231 Hz) was based on the speaker's average pitch for real laryngealized words during the first fifth of the vowel. The low F0 value (179 Hz) was selected as the lowest naturalistic F0 value that could be synthesized in Praat. The mid F0 value (205 Hz) was the middle of the two extremes, and only 5 Hz higher than the speaker's own average F0 for modally phonated words during the first fifth of the vowel. These 16 recordings were then duplicated. 30% jitter was then added to the coda sonorant of the resulting files using the *raspiness* function from the Praat Vocal Toolkit [17] to simulate the percept of creakiness or laryngealization, following Huang [18]. This manipulation decreases HNR by introducing more aperiodicity into the acoustic signal. Overall, this process produced 12 types of stimuli for a total of 96 unique stimulus items (6 F0 contours x 2 phonation conditions x 8 unique nonce words). Henceforth these recordings will be referred to as the test stimuli, or simply the stimuli.

Each task consisted of 192 test trials. During the ABX task, for each trial listeners first heard the same nonce word twice, once with naturally produced laryngealization and once with modal phonation, with the order counter-balanced across the experiment. Listeners then heard one of the test stimuli and were instructed to indicate whether they thought the test stimulus sounded more like the modal or laryngealized word by pressing a button on a button box. During the rhyming task, listeners heard a real word with either naturally produced laryngealization or modal phonation before hearing a test stimulus. They were then instructed to indicate whether the two words rhymed by pressing a button on the button box.

The results of the ABX task were coded in binary based on whether a test stimulus was matched to a laryngealized nonce word (1) or not (0). A similar coding scheme was used for trials of the rhyming task in which listeners were asked to rhyme the test stimuli with real, laryngealized words. Trials in which listeners rhymed test stimuli to real, modally phonated words were coded separately based on whether a test stimulus was matched to a real, modally phonated word (1) or not (0). Three logistic mixed effects models were then conducted in R using the *lme4* package [19] with matches to laryngealized/modal words as the dependent variable and the 12 test stimuli types as a sum-coded fixed effect. Listener, trial, and nonce word were included as random intercepts, and an additional binary fixed effect was included in the ABX model to account for a response bias in favour of the second word. In addition, to determine which types of stimuli promoted more 'laryngealized' responses than chance (50%), binomial tests were conducted on the stimuli types that promoted significantly more 'laryngealized' responses on either task using the *binom.test()* function in R.

3. RESULTS

As shown in Figure 1 and Table 2, for the ABX task, listeners were more likely to match HHˀ ('ˀ' here indicating added raspiness), HLˀ, HM, and HMˀ test stimuli to laryngealized nonce words compared to other stimulus types. This is indicated by the black asterisks in Figure 1 and grey shading in Table 2. LL, LLˀ, and MM stimuli were less likely to be matched to laryngealized nonce words, indicated by the grey asterisks in Figure 1.

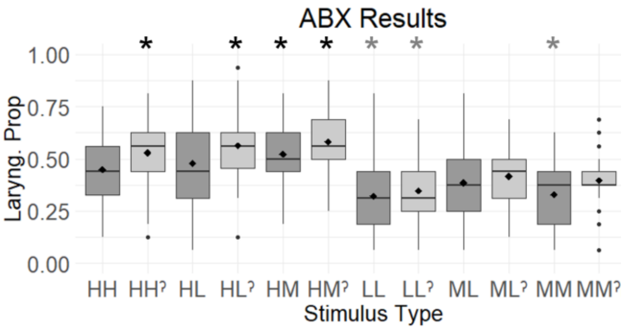

Figure 1: Proportion of 'laryngealized' responses for the ABX task. Box shading indicates whether the stimulus type had added raspiness (light grey) or not (dark grey). Asterisk color indicates whether the stimulus type promoted significantly more 'laryngealized' responses (black) or less (grey).

On the rhyming task, listeners were more likely to rhyme HLˀ, HM, and HMˀ words and less likely to rhyme LL, LLˀ, ML, and MM words to laryngealized words, as shown in Figure 2 and Table 2.

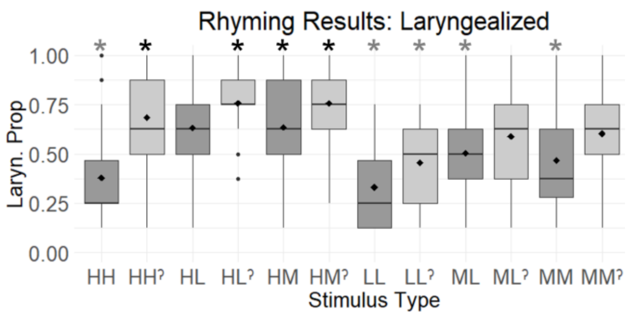

Figure 2: Proportion of 'laryngealized' responses for the rhyming task. Box shading indicates whether the stimulus type had added raspiness (light grey) or not (dark grey). Asterisk color indicates whether the stimulus type promoted significantly more 'laryngealized' responses (black) or less (grey).

Stimuli	Task	Results
HH	ABX	$\beta = 0.05$, $z = 0.226$, $p = 0.821$
	Rhyme	$\beta = -0.95$, $z = -5.293$, $p < 0.001$*
HHˀ	ABX	$\beta = 0.43$, $z = 2.056$, $p = 0.040$*
	Rhyme	$\beta = 0.55$, $z = 3.046$, $p = 0.002$*
HL	ABX	$\beta = 0.22$, $z = 1.036$, $p = 0.300$
	Rhyme	$\beta = 0.32$, $z = 1.850$, $p = 0.064$
HLˀ	ABX	$\beta = 0.61$, $z = 2.939$, $p = 0.003$*
	Rhyme	$\beta = 1.13$, $z = 6.010$, $p < 0.001$*
HM	ABX	$\beta = 0.42$, $z = 2.021$, $p = 0.043$*
	Rhyme	$\beta = 0.44$, $z = 2.486$, $p = 0.013$*
HMˀ	ABX	$\beta = 0.72$, $z = 3.449$, $p < 0.001$*
	Rhyme	$\beta = 1.13$, $z = 6.023$, $p < 0.001$*
LL	ABX	$\beta = -0.65$, $z = -3.047$, $p = 0.002$*
	Rhyme	$\beta = -1.43$, $z = -7.583$, $p < 0.001$*
LLˀ	ABX	$\beta = -0.54$, $z = -2.556$, $p = 0.011$*
	Rhyme	$\beta = -0.54$, $z = -3.009$, $p = 0.002$*
ML	ABX	$\beta = -0.29$, $z = -1.401$, $p = 0.161$
	Rhyme	$\beta = -0.41$, $z = -2.383$, $p = 0.017$*
MLˀ	ABX	$\beta = -0.12$, $z = -.0513$, $p = 0.608$
	Rhyme	$\beta = 0.22$, $z = 1.234$, $p = 0.217$
MM	ABX	$\beta = -0.61$, $z = -2.908$, $p = 0.003$*
	Rhyme	$\beta = -0.70$, $z = -4.037$, $p < 0.001$*
MMˀ	ABX	$\beta = -0.28$, $z = -1.357$, $p = 0.175$
	Rhyme	$\beta = 0.26$, $z = 1.517$, $p = 0.129$

Table 2: Logistic regression results for the ABX task and laryngealized trials of the rhyming task. Shaded rows indicate stimulus types with significantly more 'laryngealized' responses.

The results of the statistical analysis on the modal match data from the rhyming task were similar to that of the other two models. Listeners were significantly less likely to rhyme HL, HLˀ, HMˀ, and MLˀ stimuli to real, modally phonated words than other test stimulus types. Conversely, HH, LL, LLˀ, MM, and MMˀ stimuli were significantly more likely to be rhymed with real, modally phonated words. These results are shown in Figure 3 and Table 3.

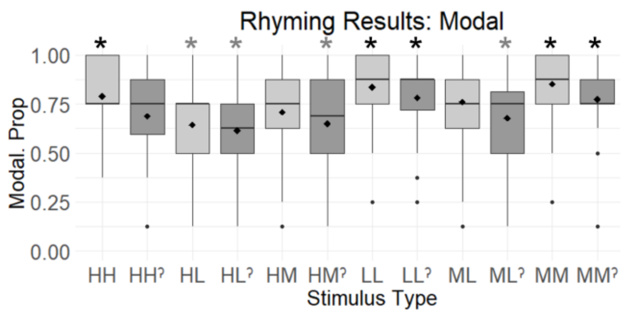

Figure 3: Proportion of 'modal' responses for the rhyming task. Box shading indicates whether the stimulus type had added raspiness (light grey) or not (dark grey). Asterisk color indicates whether the stimulus type promoted significantly more 'modal' responses (black) or less (grey).

Stimuli	Rhyming: Laryngealized Trials
HH	β = 0.47, *z* = 2.968, *p = 0.003**
HHˀ	β = -0.16, *z* = -1.107, *p = 0.268*
HL	β = -0.55, *z* = -4.195, *p < 0.001**
HLˀ	β = -0.69, *z* = -5.352, *p < 0.001**
HM	β = -0.21, *z* = -1.529, *p = 0.126*
HMˀ	β = -0.51, *z* = -3.903, *p < 0.001**
LL	β = 0.82, *z* = 4.799, *p < 0.001**
LLˀ	β = 0.41, *z* = 2.657, *p = 0.008**
ML	β = 0.08, *z* = 0.587, *p = 0.558*
MLˀ	β = -0.27, *z* = -1.978, *p = 0.048**
MM	β = 0.92, *z* = 5.354, *p < 0.001**
MMˀ	β = 0.31, *z* = 2.103, *p = 0.003**

Table 3: Logistic regression results for the modal trials of the rhyming task. Shaded rows indicate stimulus types with significantly more 'modal' responses.

Finally, binomial tests were performed for each test stimulus type which promoted significantly more 'laryngealized' responses than the overall average on either task (HHˀ, HLˀ, HM, HMˀ stimuli). The results show that on the ABX tasks, HLˀ and HMˀ test stimuli promoted significantly more 'laryngealized' responses than chance. All stimulus types tested promoted more 'laryngealized' responses than chance on the rhyming task. These results are summarized in Table 4, along with the percentage of 'laryngealized' responses per stimulus type.

Stimuli	ABX	Rhyming
HHˀ	53.5%, *p* = 0.109	68.4%, *p* < 0.001*
HL	47.9%, *p* = 0.155	64.1%, *p* < 0.001*
HLˀ	56.3%, *p* = 0.001*	77.0%, *p* < 0.001*
HM	52.3%, *p* = 0.137	65.8%, *p* < 0.001*
HMˀ	58.2%, *p* < 0.001*	77.6%, *p* < 0.001*

Table 4: Binomial test *p* values and percentage of 'laryngealized' responses per stimulus type by task.

4. DISCUSSION

The results of this study reveal three significant findings. First, listeners only perceived nonce words as 'laryngealized' when they began with a high F0 vowel, regardless of whether there were other acoustic cues to laryngealization in the stimuli (e.g., low F0 or low HNR on the coda sonorant). This suggests that there is a hierarchy between the acoustic cues to laryngealization, with high F0 on the vowel at the top.

Second, listeners only perceived words as 'laryngealized' if there were at least two cues correlating with the unique acoustics of Danish laryngealization. That is, listeners were more likely to perceive words as 'laryngealized,' compared to other test stimulus types, only if the word had both high F0 on the vowel and either a drop in F0 or low HNR (raspiness) on the coda sonorant. This result indicates that listeners utilized multiple, redundant cues to laryngealization in perception.

Furthermore, listeners were able to utilize either low F0 or low HNR as a cue to the second phase of laryngealization in Danish, which is often associated with creaky or compressed phonation. This indicates that listeners' perception of the second phase of laryngealization does not hinge on a single, specific acoustic cue. Rather, listeners may exploit various acoustic cues associated with the non-modal portion of the word to perceive the word as 'laryngealized.' This matches with previous descriptions in the Danish literature, which describe the production of the second phase of laryngealized syllables as extremely variable, acoustically [2, 14].

Finally, the more acoustic cues to laryngealization that were present, the more likely listeners were to identify test stimuli as 'laryngealized,' given the first two generalizations. That is, HLˀ and HMˀ words, which included three acoustic cues to laryngealization (high F0 on the vowel, an F0 drop, and low HNR on the coda sonorant) promoted the most 'laryngealized' responses on both tasks. These results were corroborated by the results of the binomial tests, which found that listeners were above chance at identifying HLˀ and HMˀ words as 'laryngealized' on both tasks.

Also of note is the high degree of agreement between the results of the two tasks. The ABX task was designed to examine which acoustic cue(s) that we tested were associated with the perception of laryngealized phonation by directly pitting modally phonated and laryngealized words against each other. The rhyming task, on the other hand, was designed to test listeners' ability to generalize their knowledge of the Danish phonation contrast to novel forms. Despite the differences between these two tasks, however, the results converged such that on both tasks the same types of test stimuli significantly promoted more 'laryngealized' response compared to other stimulus types. This indicates that Danish listeners are not only sensitive to the fine-grained acoustics of the phonation contrast, as shown by the ABX task, but that they are also able to generalize this knowledge to novel forms.

Overall, these results indicate that the perception of laryngealized phonation in Danish is complex, paralleling results of productions studies which find that the acoustic realization of laryngealization is multifaceted. Future research should investigate the role of other acoustic cues to laryngealization, such as spectral tilt, to the perception of the Danish phonation contrast.

5. REFERENCES

[1] Basbøll, H. 2005. *The phonology of Danish*. Oxford: Oxford University Press.

[2] Fischer-Jørgensen, E. 1989. Phonetic Analysis of the Stød in Standard Danish. *Phonetica*, 46, 1–59.

[3] Grønnum. N. 1998. Danish. *Journal of the International Phonetic Association*, 28, 99–105.

[4] Grønnum, N., Vazquez-Larruscaín, M., Basbøll, H. 2013. Danish Stød: Laryngealization or Tone, *Phonetica*, 70, 66–92.

[5] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29, 383-406.

[6] Garellek, M., Keating, P. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *The Journal of the Acoustical Society of America*, 41, 185–205.

[7] Esposito, C.M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, 40, 466–476.

[8] DiCanio, C.T. 2012. Coarticulation between tone and glottal consonants in Itunyoso Trique. *J. Phonetics*, 40, 162–176.

[9] Kuang, J. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, 142, 1693–1706.

[10] K Yu, K.M., Lam, H.W. 2014. The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, 136,1320–1333.

[11] Jiang, B., Clayards, M., Sonderegger, M. 2020. Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects. *Laboratory Phonology*, 11.

[12] Keating, P., Garellek, M., Kreiman, J. 2015. Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences*, 1, 2-7.

[13] Kirkedal, A.S. 2016. Danish Stød and Automatic Speech Recognition. Frederiksberg: Copenhagen Business School (CBS).

[14] Grønnum, N. Basbøll, H. 2001. Consonant length, stød and morae in standard Danish. *Phonetica*, 58, 230–253.

[15] Boersma P. Weenink, D. 2021. Praat: doing phonetics by computer [Computer program].

[16] Chai, Y. 2021. Perception of checked tones in Xiapu Min. *The Journal of the Acoustical Society of America*, 150, A309–A310.

[17] Corretge, R. 2022. Praat Vocal Toolkit. [Online]. Available: http://www.praatvocaltoolkit.com.

[18] Huang, Y. 2020. Different attributes of creaky voice distinctly affect Mandarin tonal perception. *The Journal of the Acoustical Society of America*, 147, 1441–1458.

[19] Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting Linear Mixed-Effects Models using lme4.