The Northern Vietnamese tonal system involves pitch and phonation differences\textsuperscript{1,2,3}. Since focus typically enhances specific acoustic properties of words, the present research addresses the question of the extent to which focus affects the phonation and other acoustic properties of Vietnamese tones. Specifically, we compare the manifestation of focus on two rising tones which (prescriptively) differ with respect to phonation: (i) modal s\textacuten and (ii) creaky ng\textacuten. While we found that duration and energy are the strongest correlates of focus for both tones as expected, we also – unexpectedly – found that there were two distinct groups of speakers with regard to the manifestation of the tones themselves. One group exhibited the typical creaky phonation with the ng\textacuten tone, while the other did not.

Focus is in-situ in Vietnamese\textsuperscript{4} and it has been reported to be realized with increased intensity, duration and F0 height / slope\textsuperscript{4,5,6,7}. The (limited) previous investigations of the interaction between focus and phonation show that focus substantially amplifies the pitch contours but only minimally affects phonation\textsuperscript{5}. It is difficult to interpret such data, however, due to confounds introduced by the carrier sentences which generally place the targets i) in focus position and/or ii) at the end of a sentence (i.e., “Say X tomorrow.” or “This is the X.”\textsuperscript{5}).

Our stimuli, (real) CVCVCV compounds, were presented in a series of dialogues shown in a PowerPoint presentation and recorded for later analysis. The target vowels, /a/ (orthographic “a”), /i/, and /u/, were produced with the s\textacuten and ng\textacuten tones and appeared in focus and non-focus contexts; only the first two syllables were considered, to avoid possible confounds with word-final properties. We analyzed a total of 1584 vowels produced by nine speakers from Hanoi, specifically measuring: F0, energy, duration and phonation properties (spectral tilt, CPP, and HNR). Additionally, we coded the data for the presence of creaky voice (CV) using auditory and visual information.

The CV coding yielded surprising results, as we found two distinct groups of speakers: those who use CV as expected with the ng\textacuten tone and others who do not. This was confirmed with a binary logistic regression analysis, which showed that the two groups distinguished the two tones equally well but by using different means. Specifically, the “creaky” speakers use both CV and F0, while the “modal” speakers only rely on F0 (Figures 1-2). With regard to focus, however, the two groups showed similar results: duration and energy were the main cues for focus (Figures 3-4). Thus, like previous studies\textsuperscript{5}, we do not find phonation to be significantly enhanced under focus; however, in contrast to previous studies, we also do not find enhancement of F0 to play a significant role. We suggest this may be due to differences in methodologies, measures, and tones examined, and most importantly, confounds with other phenomena in previous studies.

Finally, we briefly consider the implications of our two groups of speakers and suggest that there may be a change in progress affecting the tonal system of Vietnamese.
Figure 1. F0 contours

Figure 2. Harmonic-to-Noise Ratio

Figure 3. Duration

Figure 4. Energy

References: