



Festschrift for Jack Hoeksema

PERCEPTION OF THE PROSODIC FEATURES OF NEWSCASTER SPEECH

Byron Ahn¹, Z.L. Zhou², Emily Gasser³ and Donna Jo Napoli³

Abstract

It is frequently posited that the speech of groups with particular communicative goals constitutes cohesive prosodic genres. Here we investigate American English newscaster speech to discover whether listeners can distinguish it from non-newscaster speech based on prosody alone, and, if so, which features they use to do so and which are ignored. In a perception experiment, participants classified low-pass filtered audio clips as “newscaster” or “everyday” speech. We find that listeners were able to identify newscasters at a rate better than chance. They correctly utilized newscasters’ use of lower pitch, more intonational breaks, and little non-modal phonation, but overlooked the range of pitch accents that characterize newscaster speech. Whether this results from a coherent Newscaster genre or an interacting collection of indexical traits, our results suggest that listeners’ competence about speech registers includes sub-phonemic information about intonation, impacting the kind of data linguists must attend to when designing intonational models.

Keywords: prosody, newscaster speech, prosodic genres, speech perception

1. Princeton University

2. University of California at Los Angeles

3. Swarthmore College

Corresponding author: Donna Jo Napoli, donnaionapoli@gmail.com

Published by University of Groningen Press, Copyright © by authors

How to cite this article: Ahn, B., Zhou, Z.L., Gasser, E. & Napoli, D.J. (2024). Perception of the prosodic features of newscaster speech. *TABU Festschrift for Jack Hoeksema*. 203-218. <https://doi.org/10.21827/tabu.2023.41266>

This article is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License ([CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/))

1. Background^{1 2 3}

It is often argued that the speech of groups with particular linguistic and communicative goals constitutes prosodic genres, with speakers behaving cohesively, and perhaps distinctly (e.g., Johns-Lewis, 1986, and many since). Among proposed prosodic genres is newscaster speech (Gasser et al., 2019, and see background and citations therein). Recent work includes Ward (2019), regarding “atypical” prosody and Nissen et al. (2020) regarding pitch.

Much work on American English newscaster prosody takes as a given that newscasters sound distinct to a native speaker’s ear – an untested impression. Gasser et al. (2019) found a number of significant systematic prosodic differences that distinguish newscaster speech, summarized by points (1)–(7) in Table 1. Further analysis of that corpus allows us to add point (8). Using the findings in Table 1, the present paper explores two questions:

- 1) Can listeners reliably distinguish newscaster speech based solely on prosodic features?
- 2) Of the range of prosodic features newscasters employ (in Table 1), which do listeners use to identify them, and which go unnoticed?

Table 1: Summary of findings based on production

1	A higher number of L+H* pitch accents occurred in newscaster than volunteer clips.
2	L*+H pitch accents were used by volunteers and not by newscasters.
3	More Intonation Phrase breaks (level 4) were used in newscaster than volunteer clips.
4	Mean speech rate was slower in newscaster than volunteer clips.
5	Less variation in intensity was found in newscaster than volunteer clips.
6	Newscaster clips had lower pitch minima for both female and male speakers and lower pitch maxima for female speakers, despite a lack of differences in pitch averages and ranges. Male speakers did not differ in absolute pitch maxima between the two conditions.
7	Newscaster clips made most frequent use of the middle two quartiles of their pitch range; volunteer clips made most frequent use of the lowest quartile. Newscaster clips also spent similarly low amounts of time in the lowest and highest quartiles, while volunteer clips spent a relatively large amount of time in the lowest quartile and very little time in the highest quartile.

- 8 Volunteer clips had a higher incidence of non-modal voicing, operationalized as ‘undefined’ f0 measurements.
-

2. What newscasters try to do through prosody; why context matters

Newscasters are trained to achieve certain goals with their on-air voice by using slower speech rate, lower pitch, and frequent pauses. Decreased speaking rate and frequent pausing contribute to what Keerstock and Smiljanic (2019) call “clear speech”, making the content easier for listeners to recall. In a study of radio employers’, educators’, and trainers’ perceptions of successful attributes of radio voices in Australia, respondents listed vocal characteristics: “deeper, more assured” voices, “slower, more educated diction delivery” (perceived as evidence of credibility), voices that sound “real and natural”, where “warmth in a voice makes a person sound confident and honest” (Warhurst et al., 2013 p 222).

However, adopting a newscaster voice in practice appears to be automatic and largely unintentional. One news anchor, Jeannette Reyes, conveyed this in an online post (Instagram, 28 January 2021: [instagram.com/tv/CKm380kpw2H](https://www.instagram.com/tv/CKm380kpw2H)): “When I’m relaying important information at the desk, [the newscaster voice] is subconscious. I go into the anchor voice, because for me, I feel more authoritative. I automatically want to speak more clearly. My mind just goes to that voice when it’s important stuff.”

Gasser et al. (2019) find a similar discrepancy: American radio newscasters in their survey say they aim to convey objectivity, authority, and credibility; but also that their on-air voice is no different from their normal speaking voice. At the same time, the authors find that on-air newscaster speech prosody is systematically different from non-newscaster speech in the ways described in Table 1, supporting the assertion that the switch to newscaster voice is subconscious. Because few prosodic properties associated in the literature with these newscasters’ stated goals are found in their corpus, they suggest that newscasters aim to hit a balance between sounding friendly/engaging on the one hand, and credible/authoritative on the other; thus, their prosody reflects interaction of mutually opposing goals. These results support the existence of a newscaster register of which speakers have a mental model.

3. Perception study design

In a large-scale rating task, participants were asked to indicate whether they believed particular audio recordings came from a news broadcast or “everyday” speech. This study was approved by the Swarthmore College IRB committee (Protocol #14-15-45) and participants gave their informed consent before beginning.

3.1. Stimuli

The text and audio of twelve target sentences was extracted from six newscasts aired in the 1990s on WBUR, a National Public Radio affiliate in Boston, and archived in the Boston University Radio News (BURN) Corpus (Ostendorf et al., 1996). Target sentences were 11–24 words each; all were declarative statements and none were transitions to other stories or speakers. Half came from male news readers, half from female. (Target sentences appear in Appendix A.) The audio extracted from the original news broadcasts comprised the first of our three experimental stimuli types, the Newscaster condition.

To create the remaining two stimuli types, additional audio recordings were made of these same target sentences in varying contexts with volunteer (non-newscaster) speakers. These volunteers produced the target sentences embedded in a larger script: the original script from the newscast or a modified (conversational) script. These recordings yielded the Volunteer-Original script and Volunteer-Modified script conditions. No reader was aware that any particular sentence would be isolated for later analysis. The text was identical across all three recording conditions; only the type of reader (newscaster vs. non-newscaster volunteer) and discourse context (original news vs modified conversational scripts) differed.

The newscaster clips were read by professional newscasters in the BURN Corpus as part of their regular broadcasts; the other clips were read by twelve recruited volunteers: six men, six women, matched by gender to the original newscaster reader of each clip, 25–60 years old, native speakers of American English, college-educated, living in the Philadelphia area. Nine volunteers were white; two, Black; one, of South Asian heritage.

The only recordings in the newscaster condition were the original broadcasts; thus, the newscaster recordings inhabit only one discourse context. For the other conditions, speakers recorded the target sentence couched in one of two contexts. The Volunteer-Original condition recordings used the original news script as context. The Volunteer-Modified condition

embedded target sentences in conversational-sounding contexts to minimize participant awareness that they were reading sentences originally from a newscast.¹

Recordings for the two volunteer conditions were made in sound-isolated booths. Volunteers were instructed to speak “as naturally and conversationally as possible.” None had journalistic experience nor prior knowledge of the goals of the experiment. Volunteers read through each script twice, and recordings with disfluencies or diversions from the script were excluded. In three female clips small disfluencies were edited out to obtain enough usable samples; these clips performed indistinguishably from unedited clips.

The 12 target sentences in the three conditions yielded 36 recordings. Details of recording (equipment used, file compression, etc.) differed, perhaps enough for participants to differentiate recordings on the basis of recording quality (e.g., Guastavino et al., 2005), so all clips were manipulated in a four-step process to assure similar sound quality. The volunteer clips were downsampled to 16,000Hz and converted from stereo to mono to match the Newscaster-Original clips (recorded in the 1990s). Using Praat’s “Scale Intensity” function, clips were scaled to 74dB. A Hann band filter with a minimum f_0 of 80Hz, a maximum of 4,000Hz, and smoothing of 10Hz was applied to all clips to remove low background noise and the higher frequencies often missing in older recordings. Clips differed in distributed background noise, so Praat’s “Remove Noise” filter was applied to all with a filter frequency range of 80Hz-4,000Hz, and a smoothing bandwidth of 40Hz. The recordings were then low-pass filtered using a Hann band filter with a minimum f_0 of 0Hz, a maximum of 400Hz, and smoothing set to 50Hz. This removed segmental information, leaving listeners to rely solely on prosody for their judgements.

The 36 sound files² were divided into two sets of 18, one clip from each speaker in each set. To limit the experiment’s length, most participants heard only one of the sets.

3.2. Methods

The task was implemented as a Qualtrics survey online via Amazon Mechanical Turk (AMT). Participants were instructed to do the task in a quiet place with headphones, and were played a test sound clip to calibrate volume. They were told they could replay a recording as many times as they wanted before answering, but they could not return to earlier decisions for revision, nor could they progress to the next question before answering the present one.

Participants chose between “newscaster’s report” and “everyday speech” designations for each recording. To compare across clips, we calculated the percentage of responses

categorizing that recording as coming from a newscaster. Participants also indicated how confident they felt for each determination, using a Likert scale from 1 (low) to 5 (high) (as shown in Appendix B.) Two catch questions were also included to test attention.

3.3. Participants

Our data came from 481 respondents with US IP addresses who successfully completed the task. A self-reporting survey at the end gathered participant demographics. Ages ranged from 18 to 79; mean age, 37.7 years. 45.9% of individuals identified as male or men, 53.8% as female or women, and one reported gender as “Other: don’t want to say”. 97.7% were self-reported native English speakers. Six reported hearing loss. 246 completed the first stimulus set, 211 completed the second, and, due to a timing error, 24 participants completed both.

4. Results

On average, participants accurately determined whether a given clip was read by a newscaster at a rate better than chance (1-sample t ; $p < 0.001$). The distribution of participant accuracy was approximately normal (Anderson-Darling; $p < 0.005$) with an average accuracy of 57.9%, standard deviation of 11.13%, and an accuracy range from 27.8% to 88.9%.

Given that the same twelve target sentences were recorded in all three conditions, clips were filtered to obscure differences in recording quality, and the low-pass filter obscured acoustic cues other than prosody, we conclude that prosodic characteristics alone drive participants’ classifications. Further, we compared our experimental clips to sentences from WBUR broadcasts in 2018 and matched to the original set in speaker gender and general topic area. Of the prosodic features considered here, these differed significantly from our experimental set only in variability of intensity, pitch maximum for male speakers, and pitch range. Of these, only intensity variability is a significant predictor of a clip’s Newscaster rating, suggesting that NPR newscaster speech at least has stayed relatively stable over time, and that the age of our newscaster clips compared to the volunteer recordings does not play a determining role in how speakers classify them.

Participants were most accurate in their classification of newscaster clips. Accuracy scores for each condition (Newscaster, Volunteer-Original script, Volunteer-Modified script)

are reported in Table 2. Between groups, accuracy scores were significantly different for all pairings (Table 3).

Table 2: Rate of correct classification as “newscaster” vs. “everyday speech”

	Newscaster	Volunteer-Original	Volunteer-Modified	Overall
Accuracy rate	67.4%	61.1%	45.2%	57.9%

Table 3: Two-sample t-test for correct classification of “newscaster” vs. not

	Newscaster~ Volunteer-Original	Newscaster~ Volunteer-Modified	Volunteer-Original~ Volunteer-Modified
two-sample t-test	$t(6058)=5.13 p<0.001$	$t(6058)=17.85 p<0.001$	$t(6058)=12.54 p<0.001$

Listeners were better at positively identifying newscaster speech than at excluding other speech from the newscaster category, as reflected in patterns of recall and precision. Average recall was 0.674; precision, 0.419; and F-score, 0.517.³ Since only one third of the clips came from newscaster speech, participants may have been looking for a more even distribution, thereby influencing their behavior. We posit further that positive identification of newscaster speech, a genre, is easier than identification of “everyday speech,” which encompasses multiple genres; consistent with this, participants were most confident in the newscaster-original condition (Table 4), and the difference in confidence between newscaster and volunteer conditions is significant, while there was no significant difference between the two volunteer conditions (Table 5).

Table 4: Confidence in classification of “newscaster” vs. not

	Newscaster	Volunteer-Original	Volunteer-Modified	OA
mean confidence	3.68 /5	3.54/5	3.54/5	3.59/5

Table 5: Two-sample t-test for confidence in classification of “newscaster” vs. not

	Newscaster~ Volunteer-Original	Newscaster~ Volunteer-Modified	Volunteer-Original~ Volunteer-Modified
two-sample t-test	$t(6058)=5.038 p<0.001$	$t(6058)=5.273 p<0.001$	$t(6058)=0.187 p=0.852$

Figure 1 shows the (nonsignificant) interaction of confidence and Newscaster ratings for newscaster clips versus volunteer clips.

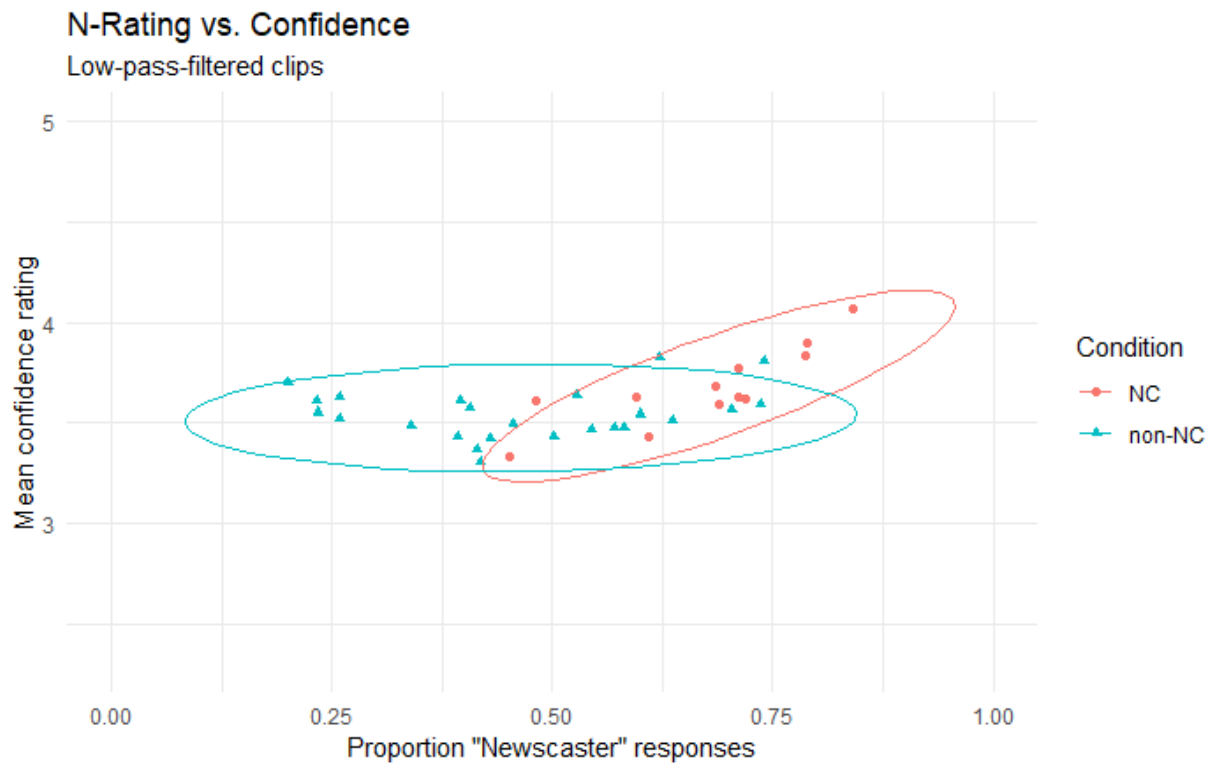


Figure 1. Each dot represents the proportion of participants responding “newscaster” (noted as N-rating here) and average confidence rating for each clip; ellipses show the 95% confidence interval. NC denotes Newscaster clips; non-NC denotes non-Newscaster (i.e., Volunteer) clips.

Clips with the highest average confidence ratings also had the highest Newscaster ratings. While there is overlap between how newscaster and volunteer clips were rated, as groups they exhibit separate patterns.

We tested for statistical correlation between a participant’s overall accuracy (percentage of responses that correctly categorized a clip as Newscaster) and various measures. No demographic information exhibited significant influence, nor did participant confidence in their identification correlate with increased identification accuracy ($p > 0.05$).

Figure 2 presents plots for the proportion of “newscaster” responses for all clips, grouped by recording condition. The three conditions yielded significantly different responses.

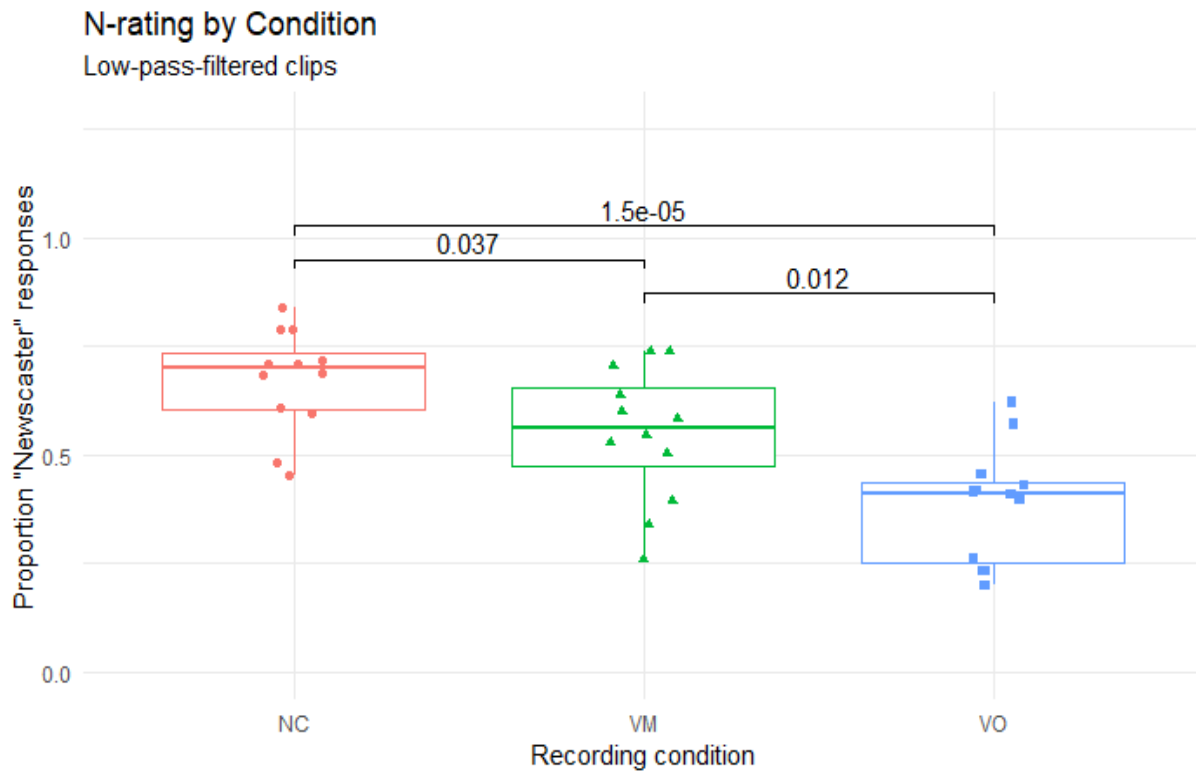


Figure 2. Percentage of responses deeming the recording as being a newscaster's. Each dot represents the proportion of participants responding “newscaster” (noted as “N-rating” here) for a particular clip, and horizontal lines show the mean of those responses. NC denotes the original newscaster clips, VM the Volunteer-Modified script condition, and VO the Volunteer-Original script condition.

The average proportion of “newscaster” responses was highest for the newscaster clips (Newscaster-rating_{NC}=.674); volunteer clips had average proportions of “newscaster” responses around or below 50% (Newscaster rating_{VolunteerModified}=.548, Newscaster rating_{VolunteerOriginal}=.389). To investigate which cues influenced categorization, we modeled our results with a mixed effects logistic regression with the lme4 package in R. Speaker and participant were included as random variables; clip item and random slopes were not.

First, we created a maximal model: This included ToBI-labeled features (pitch accents and phrasal disjunctures) plus the other variables which differ significantly between groups in the sample (see Table 1), listed in Table 6. We used clip length as a proxy for speech rate.

Table 6: Variables in maximal model

ToBI features	number of {H*, L*, !H*, L+H*, L*+H}; number of {ip breaks, IP breaks}
acoustic features	normalized pitch minimum, normalized pitch maximum, usage of different pitch quartiles, proportion of undefined pitch, clip length, variability of intensity

Next, we conducted model selection. Starting from a maximal model, we created a series of subset models, each lacking the least significant predictor of the previous one. This procedure continued until we arrived at a model with every predictor significant at the $\alpha=0.05$ level. These models were then compared using the `anova()` function. Table 7 presents the model that best predicted the data with a minimum number of predictors ($\chi^2=48.97, p\approx 0$).

Table 7: Parameter values for best fit model

	Estimate	Std. Error	z value	Pr(> z)	
<i>(Intercept)</i>	-2.288	0.367	-6.233	4.58E-10	***
# of H*'s	-0.049	0.029	-1.684	0.092	.
# of L*'s	-0.322	0.046	-6.99	2.74E-12	***
# of ip's	0.125	0.037	3.348	0.001	***
# of IP's	0.542	0.101	5.365	8.11E-08	***
length	0.233	0.054	4.319	1.57E-05	***
intensity range (db)	0.010	0.004	2.598	0.009	**
pitch maximum (z-scored)	-0.187	0.067	-2.786	0.005	**
% of f0 in lowest quartile	1.349	0.221	6.102	1.05E-09	**
% of clip with 'undefined' f0	-1.356	0.743	-1.825	0.068	.

Participants were more likely to categorize a clip as newscaster the more prosodic phrases it included, both intermediate phrase and Intonation Phrase, and less likely to do so the more H* and L* pitch accents occurred. They were more likely to categorize a clip as newscaster if it is longer, has a greater intensity range, and if more time is spent in the lowest quartile of its pitch

range, and less likely if it has a higher pitch maximum and a larger proportion of ‘undefined’ f0. Overall, participants seemed to believe that newscasters use fewer simplex pitch accents, more intonational breaks, and speak more slowly and with more variable loudness than non-newscasters. Additionally, participants seemed to believe that newscasters speak with a lower pitch but little non-modal phonation.

Based on previous findings (see Table 1), these effects can be split into three categories. Listeners were accurate in their use of the number of Intonation Phrases, proportion of non-modal phonation, and speech rate to categorize clips: These features do differentiate newscaster from non-newscaster speech, and in the direction reflected by participants’ responses. Listeners appeared to use the number of simplex H* and L* pitch accents and intermediate phrases in their categorization of clips, even though they did not in fact differ between newscaster and volunteer recordings. Listeners were misled by intensity variability and time spent in the lowest pitch quartile: though in fact newscasters had lower variability of intensity and spent less time in their lowest pitch quartiles than non-newscasters, these features had the opposite effect on participants categorizations. Listeners were more likely to categorize a clip as newscaster if it had a lower pitch maximum; in fact this differed between categories for female speakers but not male speakers. Additionally, listeners ignored several features which are in fact informative of newscaster status: a higher number of L+H* pitch accents, lack of L*+H pitch accents, lower minimum F0, and more time spend in the middle 50% of their pitch range.

5. Discussion and conclusion

The answer to our Question 1: Participants were able to identify newscaster speech based solely on prosodic cues. No demographic features had a significant effect on accuracy, nor did reported confidence. The lack of demographic effects suggests that our results are generalizable across the population of American English speakers. This study is consistent with other research showing that speakers have complex cognitive representations for the speech of other groups (Wade, 2020). In addition to the significant effects stemming from categorical intonational factors such as frequency and identity of pitch accents and phrase boundaries, we also found effects of continuous variables such as speech rate. While the categorical variables we investigated in this study are straightforwardly linguistic in nature, the continuous ones are rather less so. In general, it could be said that newscaster and non-newscaster speakers use the same phonological intonational categories (perhaps barring L*+H), but implement them

differently in the phonetics. This suggests that both producers and perceivers have a mental model of a speech style that may contain socially-oriented and/or sub-phonemic information.

Regarding Question 2, the newscasters in our corpus use lower pitch and speak more slowly, with little non-modal phonation and more intonational breaks; listeners were more likely to categorize clips with these features as coming from a newscaster. However, the complex pitch accents characteristic of newscaster speech, as well as tendency to stay in the middle 50% of their pitch range, did not influence listeners' categorization of clips – though this information was informative, they overlooked it in their judgements, and it therefore does not seem to contribute to the mental model of newscaster speech.

The Newscaster rating results are surprising in that the Volunteer-Modified script condition was categorized more frequently as “newscaster” than the Volunteer-Original script condition with its original newscaster script. This is especially puzzling as volunteers who read for the Volunteer-Original script condition reported that they could tell the text originated from a news broadcast and involuntarily fell into a “newscaster-y voice”. One explanation is that the Volunteer-Original scripts demanded of the readers that they do something unnatural: read the script in a style that was at odds with its content. In contrast, the Newscaster condition had professional newscasters reading newscaster scripts as newscasters, and the Volunteer-Modified script condition had non-newscasters reading conversational scripts conversationally. In this respect, perhaps listeners were attending to some markers of fluency or comfort, and readers in the Volunteer-Original script condition portrayed these features the least. A methodological lesson here is that asking participants to portray the speech of a register they are unpracticed in may lead to unusual results.

It is also possible that there is in fact no newscaster register *per se* and that listeners are, instead, listening for cues indexing objectivity, authoritativeness, and other related traits, using these as a proxy for “newscaster”. This possibility would be consistent with Gasser et al.'s (2019) finding that newscasters try to promote an objective and dispassionate stance, aiming for credibility and trustworthiness. Given that listeners associate paralinguistic disfluencies such as those discussed above with deception (Loy et al., 2016, 2017), the Volunteer-Original clips might well sound like bad liars. If our participants believe that newscasters will be honest, then bad liars are precluded.

This hypothesis would also explain participants' lackluster accuracy in categorization, since the prosodic features used by newscasters have been shown to map imperfectly onto the traits that they aim to convey (Gasser et al., 2019 and references therein). However, this fails to fully explain the cohesiveness shown by the newscasters in their use of the register. We

believe it is likely that the newscaster prosodic register arose from the convergence of the features associated with key pragmatic, sociological, and paralinguistic features, but has since been abstracted somewhat from those roots and is now an imitative phenomenon.

There are broad implications from this study. The differing results of Volunteer-Modified vs. Volunteer-Original clips confirms that context informs prosody (Beun, 1990) in ways listeners attend to (Vaissière, 2005). Since our clips are read-speech, which differs prosodically from spontaneous speech (Ayers, 1994), context plays a role even in reading tasks, (see Cole, 2015): context must be supplied in read-speech tasks to get reliable results.

That listeners' competence about speech registers includes sub-phonemic information about intonation impacts what kind of data linguists must attend to when designing an intonational model for a language. Such a model must include speech from multiple types of contexts, where corpus data are tagged for context/discourse-mode, allowing the model to deploy the appropriate intonation for the context. The methodology applied here can serve for identifying the prosodic features that coincide with any number of other groups. Thus, we might look closely at doctor-patient, teacher-pupil, parent-child discussion, and so on.

Endnotes

¹ Original and modified scripts can be found at <https://doi.org/10.5281/zenodo.8350354>

² Low-pass filtered sound files are available at <https://zenodo.org/record/8361174>. Unfiltered sound clips are at <https://zenodo.org/record/8361211>

³ To calculate recall/sensitivity, divide the number of clips participants correctly classified as newscaster by the actual number of newscaster clips (rate of true positives). To calculate precision, divide the number of clips participants correctly classified as newscaster by the total number that participants (correctly or incorrectly) classified as newscaster. F-score is the harmonic mean of precision and recall.

References

- Ayers, G. M. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics*, 44, 1–49.
- Beun, R-J. (1990). The recognition of Dutch declarative questions. *Journal of Pragmatics*, 14(1), 39–56.
- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, 30(1–2), 1–31.

- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23.
- Gasser, E., Ahn, B., Napoli, D. J., & Zhou, Z. L. (2019). Production, perception, and communicative goals of American newscaster speech. *Language in Society*, 48(2), 233–259.
- Guastavino, C., Katz, B. F. G., Polack, J.-D., Levitin, D. J., & Dubois, D. (2005). Ecological validity of soundscape reproduction. *Acta Acustica United with Acustica*, 91, 333–341.
- Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 199–220). College-Hill Press.
- Keerstock, S., & Smiljanic, R. (2019). Clear speech improves listeners' recall. *Journal of the Acoustic Society of America*, 146(6), 4604–4610.
- Loy, J. E., Rohde, H., & Corley, M. (2016). Lying, in a manner of speaking. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedings of Speech Prosody 8* (pp. 984–988). Boston University.
- Loy, J. E., Rohde, H., & Corley, M. (2017). Effects of disfluency in online interpretation of deception. *Cognitive Science*, 41, 1434–1456.
- Nissen, S., Randle, Q. B., Johnson, J. L., & Lynes, J. (2020). Prosodic elements for content delivery in broadcast journalism: A quantitative study of vocal pitch. *Electronic News*, 14(2), 63–77.
- Ostendorf, M., Price, P., & Shattuck-Hufnagel, S. (1996). Boston University radio speech corpus, LDC96S36. Linguistic Data Consortium.
- Pierrehumbert, J. B., & Hirschberg, J. (1990) The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271–312). MIT Press.
- Vaissière, J. (2005). Perception of intonation. In D. Pisoni & R. Remez (Eds.), *The Handbook of Speech Perception* (pp. 236–263). Blackwell.
- Wade, L. (2020). The Linguistic and the Social Intertwined: Linguistic Convergence Toward Southern Speech. Doctoral dissertation, University of Pennsylvania.
- Ward, N. G. (2019). *Prosodic Patterns in English Conversation*. Cambridge University Press.
- Ward, N. G., Carlson, J. C., & Fuentes, O. (2018). Inferring stance in news broadcasts from prosodic-feature configurations. *Computer Speech & Language*, 50, 85–104.
- Warhurst, S., McCabe, P., & Madill, C. (2013). What makes a good voice for radio: Perceptions of radio employers and educators. *Journal of Voice*, 27(2), 217–224.

Appendix A: Complete list of target sentences


Below are the twelve target sentences used as the text for recordings made in all three conditions. The code in parentheses indicates the script number, as found in the Boston University Radio News Corpus (Ostendorf et al., 1996).

1. Price was making his third start for Boston since he was signed as a free agent last month. (f1as30p5)
2. The Red Sox beat the first place Baltimore Orioles five to three this afternoon at Fenway Park. (f1as41p6)
3. Grilsh says he's a product of the hearing world and it's frustrating to no longer be able to participate fully. (f2bs30p1)
4. Grilsh hasn't learned sign language because everyone he knows can hear. (f2bs30p1)
5. You've never seen or heard of the victim but you know the punishment is death in the electric chair. (f3asx4p1)
6. Randall Adams spent twelve years in prison before Texas finally overturned his conviction two years ago. (f3asx4p1)
7. Hack is studying the effect these sounds could have on insects which can hear the noises. (m3bs02p4)
8. No one is sure how the insects figure out which trees are withering. (m3bs02p4)
9. And his administration has not exactly welcomed the parking tax proposal either. (m4bs60p6)
10. But the T apparently knows that parking is a lucrative source of income. (m4bs60p6)
11. The legislature authorized a four hundred twenty-million-dollar reduction in Medicaid's account but left it to Weld to decide which services must go. (m4bs62p1)
12. Weld has also warned that he'd veto any changes to local property tax laws which do not allow for a voter referendum. (m4bs62p1)

Sentences (1)–(6) were originally recorded by three female newscasters on WBUR (indicated by the f1, f2, and f3 at the start of the information within parentheses following each sentence), and were consequently also recorded by female volunteer readers for this study. Similarly, sentences (7)–(12) were originally recorded by two male newscasters on WBUR, and consequently also recorded by male volunteer readers. Matching the sex of the speaker within the target sentence in this way was done so as to avoid any confound of the effects of sex on participants' perceptions.

Appendix B: Qualtrics interface for experiment

Does this recording sound like a newscaster's report, or everyday speech?



Newscaster

Everyday speech

< >

How confident are you?

On a scale of 1-5, with 1 being not confident at all, and 5 being very confident.

1 – Not at all

2

3 – Neutral

4

5 – Very confident

< >

Figure B1. The Qualtrics survey screen presented with each recording