# The Acoustic Characteristics of *um* and *uh* in spontaneous Canadian English

*Gabrielle Morin, Benjamin V. Tucker*
*Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada*

## Abstract

*The present study investigates and compares the acoustic characteristics of* uh [ə] *and* um [əm] *spontaneous speech. The data comes from a corpus of Western Canadian conversational spontaneous speech. Measures of duration, fundamental frequency, F1 and F2 were extracted from 1,048 instances of* um *and* uh. *Results indicate that longer durations occurred when markers preceded silent pauses.* Um *was found to have higher F1 and lower F2 than* uh. *F0 was overall lower for* um *in comparison to* uh. *These results provide a preliminary understanding of* um *and* uh *as markers in spontaneous Canadian English. Canadian English shows a similar proportion of* um *over* uh *usage in comparison to American and British English. Findings on vowel duration show no significant difference between um and uh. Differences in f0, F1 and F2 provide additional indication of how* um *and* uh *are different.*

## Introduction

*Um* [əm] and *uh* [ə] have been reported to be among the most frequently observed disfluencies in spontaneous speech (Shriberg, 2001). We follow Le Grézause (2017) and classify *um* and *uh* as markers as opposed to fillers and filled pauses. In the present study, we investigate the acoustic characteristics of *um* and *uh* in Canadian English.

There are considerable differences across languages with regard to the frequency of occurrence of these markers. Over the past five decades, there has been an increase in *um* occurrences across British and American English dialects while *uh* has significantly decreased (Wieling et al., 2016). In their analysis of multiple spoken language corpora, Wieling et al. (2016) found that the proportion of *um* over *uh* increased from 0.3 to around 0.5 for female speakers of American English and British English as of 2013. They also found that the frequency of *um* occurrence relative to all other words has been consistently increasing in American English. Their results show a significant relationship between age and frequency of occurrence of *um* in all four American and British English corpora, demonstrating the tendency for younger generations to use *um* more than older generations (Wieling et al., 2016). However, Horváth (2010) found a greater usage of *uh* in comparison to *um* in Hungarian spontaneous speech. There is very little research with regard to Canadian English and the occurrence of *um* and *uh* as markers. Part of this may be because Canadian English is often combined with American and/or British English dialects rather than being examined individually. Canadian English is also interesting because it has strong historical influences from British English and currently remains in close contact with American English (Boberg, 2010).

Previous work has shown that the duration of *um* is consistently greater than *uh*, likely because it is composed of two phonemes rather than one (Clark & Fox Tree, 2002; Swerts, 1998). However, data analyzing the vowel duration alone has found that *um* is shorter than *uh* (Hughes et al., 2016). The duration of these markers plays an important role in the surrounding environments. Markers have been categorized into major (*um*) or minor (*uh*) delays depending on their following silence, where *um* tends to precede longer pauses than *uh* (Clark & Fox Tree, 2002; Swerts, 1998).

Fundamental frequency (f0) is another phonetic property that has the potential to differentiate *um* and *uh* (Shriberg, 2001). While the f0 patterns can vary depending on the surrounding environments, there is evidence that the f0 of markers is generally lower than the speaker's relative f0 levels (Gabrea & O'Shaughnessy, 2000), with *uh* having a lower f0 than *um* in Dutch (Swerts, 1998). Analyzing the formants and intensity of the vowel segments in each marker can also signal differences in the production of *um* and *uh.* Work by Hughes et al. (2016) did not find major differences between F1 and F2 for the vocalic midpoints of *uh* and *um*.

The present study investigates two main questions of interest. First, is *uh* or *um* the most common form of marker found in Canadian English speech? Second, what are the acoustic characteristics of *uh* and *um*? In order to address the second question, measures of duration, fundamental frequency, F1 and F2 were extracted for each marker. Following previous research, we hypothesize that:

1. *Um* and *uh* will have an equal occurrence frequency across speakers (Wieling et al., 2016).
2. *Uh* will have a longer vowel duration than *um* (Hughes et al., 2016).
3. *Uh* will have a lower f0 than *um* (Swerts, 1998).
4. *Um* and *uh* will have similar F1 and F2 values (Hughes et al., 2016).

## Method

### *Corpus*

The conversational speech data used in this analysis is from the *Corpus of Spontaneous Multimodal-Interactive Language* (CoSMIL) (Järvikivi & Tucker, 2015). Sixteen native Canadian English speakers (14 female and 2 male; 18-23 years old) participated in the recording sessions. Participants were undergraduate students enrolled in an introductory linguistics course at the University of Alberta, each receiving credit for their participation. Participants signed up as pairs and came to do the experiment together.

The recordings were made in an observation studio, which was set up to use two high quality head-mounted microphones and two opposing ceiling mounted video cameras. The researcher controlled data acquisition from a control room and could observe the interaction via a one-way mirror. Participant pairs engaged in a 45-minute conversation while sitting across from each other in the observation room. Each participant was fitted with an over the ear omnidirectional head-mounted microphone (Countryman E6) with a flat frequency response cap. Each speaker was recorded on one channel of a stereo recording, which were subsequently separated into individual files for each speaker for later analysis. Topics were provided to help initiate conversation, however the conversation portion of the experiment was not controlled and participants were encouraged to talk about whatever topic they wanted. As a result conversational topics varied widely between participants. All sixteen recordings from CoSMIL were time aligned with orthographic transcription and phonetically aligned for Praat (Boersma & Weenink, 2020) using the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008). These alignments were used in the analysis for this study.

### *Data Extraction*

A custom script was written to extract our acoustic measures of interest via Praat. We extracted vowel duration, mid-point formant values (F1 and F2), and mean f0 of each vowel. As a control, we also extracted speech rate, which was defined as the number of syllables produced in the surrounding 6 seconds (3 second preceding and 3 seconds following). Finally, we extracted the preceding and following word along with their duration. Data was extracted from individual speakers so that f0 and formant extraction values could be appropriately tailored to each speaker. Markers for this study were defined as *um, uh,* and *er*. For comparison purposes, we also counted the instances of *like* produced by each participant. *Like* can also be used as a marker with a range of grammatical functions in spontaneous speech of Western Canadian English (Podlubny et al., 2015).

### *Statistical Analysis*

The majority of the data was modeled using Linear Mixed Effects Regression (Bates et al., 2015) with subject as a random effect. We investigated vowel duration, F1, F2 and fundamental frequency as dependent variables for the *um* and *uh* markers. We used the identity of the Marker (*um* or *uh*), the Following or Preceding context (word vs silent period (sp)), and Speech Rate as our independent variables. We used a backward stepwise model fitting procedure testing individual predictor effects along with possible two-way interactions. Non-significant effects were removed until a final best-fit model was achieved. Effects were considered significant if the $t$ value exceeded an absolute value of 2. All possible random slopes were explored after the stepwise modeling procedure and any random slopes which improved the models fit and did not result in an error or warning were retained.

## Results

A total of 1,055 markers were extracted from the eight conversations in the CoSMIL dataset, or about 66 markers per speaker with their rate of production ranging from 20 to 129 markers per conversation. Of the markers there were 7 instances of *er*, 498 of *uh*, and 550 *um*. As a result of so few instances of *er*, these were excluded from the statistical analyses leaving 1,048 instances of *um* and *uh* markers. We also counted a total of 5,513 instances of *like* in the corpus or about 344 instances of *like* per speaker, with individual speakers ranging between 151 to 585 productions of *like* during their conversations.

### *Duration*

As an initial model, we performed a t-test to compare the duration of the Marker. In this analysis, *um* (mean 428 ms) is significantly longer ($t(918.15)=-2.836$, $p < 0.005$) than *uh* (mean =243 ms) which is likely due to the fact that *um* is made up of two segments.

We then investigated the duration of the vowels in the marker, as illustrated in Figure 1. We investigated all two-way interactions in an attempt to find the most parsimonious model. We report only those predictors and interactions that were significant in the best fitting model as described in the Statistical Analysis section. Marker by Subject as a random slope improved the model fit and was retained in the final model. There is a significant interaction between Marker and Following context. The interaction illustrates that when the Following context is held constant the Markers are not significantly different from each other (sp: *t*=-0.065; word: *t*=-1.825). When the Marker is held constant there is a significant difference as a result of the Following context. For both *uh* and *um* the vowel is shorter when the Following context is a word (*uh*: *β*=-0.0842, *se (standard error)*=0.0112, *t*=-7.487; *um*: *β*=-0.0457, *se*=0.011, *t*=-4.142). When a word follows the Marker the duration of the vowel is shorter. (*β*=-0.0805, *se*=0.0122, *t*=-6.598). We also find that the faster the speech rate the shorter the vowel (*β*=-0.009, *se*=0.003, *t*=-2.934).

### Fundamental Frequency

In our f0 data there were instances where the pitch tracking algorithm failed to extract a valid measure and these items were excluded from the analysis, leaving 1029 items for the analysis. No random slopes were found to improve model fit. We have chosen not to transform our f0 values in this model as most of our speakers are female and it is hoped that the speaker random effect will account for some of the speaker variability. We found that f0 is lower when the segment is shorter (effect size: 40Hz, *β*=-43.987, *se*=12.292, *t*=-3.578) and the f0 is lower when the speech rate is faster (effect size: 62Hz, *β*=-8.745 *se*=1.292, *t*=-6.768). The f0 is slightly higher for the *uh* markers (8Hz, *β*=8.349, *se*=3.612, *t*=2.312). The f0 is higher when there is a following word as opposed to when there is following silence (11Hz, *β*=11.738, *se*=3.256, *t*=3.605).

### Formants

We also analyzed the formant characteristics of the vowels in *um* and *uh*. This comparison is illustrated in the vowel plot in Figure 2. In this analysis we transformed the formant values using the $\log_{10}$ function and also included Segment Duration as a covariate in the model. In the model of F1 no random slopes were found to improve the model fit and in the F2 model Marker and Previous context by subject significantly improved the model fit.

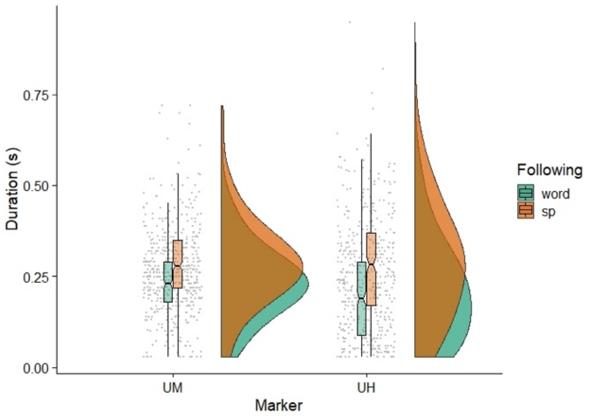

*Figure 1. Raincloud plot (Allen et al., 2021) of the vowel durations of the markers* um *and* uh *split by the following content, silent period (sp) is in brown and lexical content (word) is in green.*

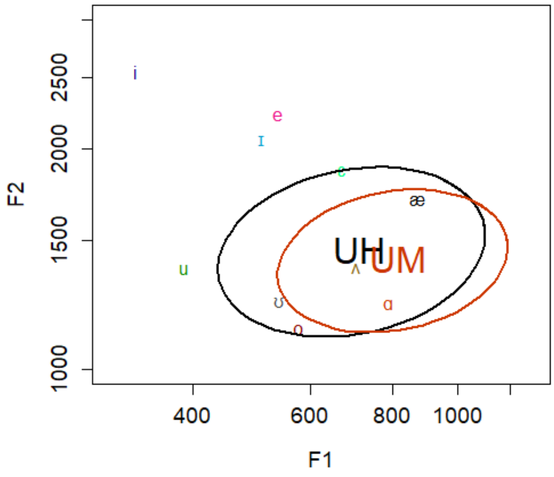

*Figure 2. Log transformed F1 by F2 plot of formant measures for the markers* um *and* uh *using phonTools (Barreda, 2015). The label indicates the average formant value and ellipses are plotted at 1.96 standard deviations. Average values from Hillenbrand et al. (1995) are plotted to provide some context.*

We found that *um* has a significantly higher F1 compared to *uh* (*β*=-0.078, *se*=0.014, *t*=-5.517) and that F1 is higher when the following item is a word (*β*=-0.028, *se*=0.012, *t*=2.203). There was also a significant interaction between Segment Duration and the Previous context. When the Previous context is a silent period there is a slight increase (effect size: 66Hz, *β*=-0.17, *se*=0.0602, *t*=2.822) in F1 as Segment Duration increases but when the preceding context is a word, we see that as segment duration increases the frequency of F1 also increases (effect size: 289Hz, *β*=0.374, *se*=0.094, *t*=3.981). As speech rate increases so does F1 frequency (*β*=0.015, *se*=0.005, *t*=2.99).

In the model analyzing F2 we find that *uh* has higher F2 than *um* (*β*=0.054, *se*=0.0145, *t*=3.726).

Speech rate was not significant in this model. Previous and Following context significantly interacted with Segment Duration: the effect is the same for both. There is no effect when the preceding context is a silent period but the effect is significant when there is a word preceding ($\beta$=-0.128, *se*=0.052, *t*=-2.453) or following ($\beta$=-0.12, *se*=0.051, *t*=-2.336). In both cases, when a word is present, longer segment duration decreases the F2.

## Discussion

In summary, the results from the present analysis indicate that there is a fairly equal but small bias toward the occurrence of *um* (550) as opposed to *uh* (498) in the 1,048 extracted markers. In testing our first research hypothesis, we find that *um* is the more common form of marker found in Canadian English speech, though only slightly more common. These results also confirm our original hypothesis that Canadian English would reflect the usage of *um* and *uh* of other English varieties, showing a similar proportion of *um* over *uh* instances as those found most recently in 2013 (Wieling et al., 2016). Following Wieling et al., (2016), we suspect that the similarity in marker proportion is likely due to cross-linguistic changes within native English speaking countries that are often influenced by societal extralinguistic forces. Interestingly, our data indicates a relatively low occurrence of *um* and *uh* markers when compared to the occurrences of *like* in the corpus. Our counts indicate that *like* occurs 5 times as often. We have not seen previous comparisons of these markers and believe that the high frequency of *like* is potentially due to the increased functional role it plays in speech (Podlubny et al., 2015).

Our findings on overall marker duration confirm that *um* has a longer duration than *uh*, likely due to the phonemic difference between the two markers (*um* /əm/ has two phonemes while *uh* /ə/ has one, Clark & Fox Tree, 2002; Swerts, 1998). These overall marker durations are consistent with previous findings. Contrary to Hughes et al. (2016) and our second research hypothesis, we do not find a significant difference in the duration of the vowels in *um* and *uh*. However, most of our participants are female, while all of the participants from Hughes et al. (2016) were male. It is possible that there are gender differences in the usage of the two markers. We do note that the reported vowel durations for both markers in our study in comparison to Hughes et al. (2016) might suggest that the vowel duration of *um* is longer in Canadian English than in other dialects. The duration of both the *uh* and *um* vowel segments are longer when followed by a silent pause than when followed by a word, suggesting that Canadian English aligns with previous claims that these prolonged vowels are used by speakers to signal an upcoming delay (Clark & Fox Tree, 2002).

The results show other acoustic phonetic differences between *um* and *uh* as well. Specifically, our third research hypothesis investigates fundamental frequency. We find that fundamental frequency is slightly lower for *um* in comparison to *uh*, disconfirming our original hypothesis (Swerts, 1998). We believe this may be due to the following voiced nasal contributing to a lower f0 in the *um* vowel, however results concerning the effect of following consonants on vowels is variable (Hanson, 2009). While the present findings generally agree with the literature, we are cautious in our interpretations as the sample size is fairly limited.

For our fourth and last research hypothesis we found that *um* has a higher F1 and lower F2 than *uh*, contradicting our original hypothesis (Hughes et al., 2016). We suspect this difference is due to the high between-speaker variability and stylistic differences that are often reported in acoustic analyses of filled pauses (Hughes et al., 2016; Clark & Fox Tree, 2002) as well as the gender differences noted previously.

We believe that additional research of Canadian English spontaneous speech datasets is necessary and recommend two possible directions. First, additional investigation of *like* as a marker in spontaneous speech is necessary. *Like* is an increasingly common marker that fills many functions in conversational speech (e.g., Fox Tree & Tomlinson Jr., 2007; Podlubny et al., 2015). Acoustic characteristics of *like* have been shown to signal its usage as a marker in comparison to its other functions (Podlubny et al., 2015). Second, further investigation of the functional role of the *um* and *uh* as stance markers (Le Grézause, 2017) in Canadian English is important. Following Swerts (1998), investigation of an interaction between phrase position, fundamental frequency, and duration for *um* and *uh* would be beneficial. The current results are an important first step to our preliminary understanding of the acoustic characteristics and differences of *um* and *uh* in spontaneous Western Canadian English.

## References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. 2021. Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4: 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Barreda, S. 2015. phonTools: Functions for phonetics in R. R package version 0.2-2.1.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Boberg, C. 2010. *The English Language in Canada : Status, History and Comparative Analysis*. Cambridge University Press.

Boersma, Paul & Weenink, David 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.32, retrieved 12 November 2020 from http://www.praat.org/

Clark, H. H., & Fox Tree, J. E. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84(1): 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Fox Tree, J. & Tomlinson Jr., J. 2007. The Rise of *Like* in Spontaneous Quotations. *Discourse Processes*, 45(1): 85-102. https://doi.org/10.1080/01638530701739280

Gabrea, M., & O'Shaughnessy, D. 2000. Detection of filled pauses in conversational speech. *ICSLP 2000*, 4.

Hanson, H. M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, 125(1): 425–441. https://doi.org/10.1121/1.3021306

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5): 3099–3111. https://doi.org/10.1121/1.411872

Horváth, V. 2010. Filled pauses in Hungarian: Their phonetic form and function. *Acta Linguistica Hungarica,* 57(2–3): 288–306.

Hughes, V., Wood, S., & Foulkes, P. 2016. Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech Language and the Law*, 23(1): 99–132. https://doi.org/10.1558/ijsll.v23i1.29874

Järvikivi, J. & Tucker, B.V. 2015. Corpus of Spontaneous Multimodal Interactive Language (CoSMIL). University of Alberta.

Le Grézause, E. 2017. *Um and Uh, and the Expression of Stance in Conversational Speech*. University of Washington.

Podlubny, R. G., Geeraert, K., & Tucker, B. V. 2015. It's All About, *Like*, Acoustics. *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–4. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0477.pdf

Shriberg, E. 2001. To 'errrr' is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1): 153–169. https://doi.org/10.1017/S0025100301001128

Swerts, M. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30(4): 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. 2016. Variation and Change in the Use of Hesitation Markers in Germanic Languages. *Language Dynamics and Change*, 6(2): 199–234. https://doi.org/10.1163/22105832-00602001

Yuan, J., & Liberman, M. 2008. Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*, 5687–5690.