IT'S ALL ABOUT, *LIKE*, ACOUSTICS

Ryan G. Podlubny[1,2]*, Kristina Geeraert*[2]*, & Benjamin V. Tucker*[2]

[1]University of Canterbury, Christchurch, New Zealand
[2]University of Alberta, Edmonton, Canada
ryan.podlubny@pg.canterbury.ac.nz; geeraert@ualberta.ca; bvtucker@ualberta.ca

ABSTRACT

The present study explores the possibility of systematic acoustic differences that could be used to differentiate 'homophones'. This study investigates productions of *like* in western Canadian English, focussing specifically on acoustic characteristics and whether they differ across multiple distinct lexical and grammatical functions. Segment duration, word duration, and degree of diphthongization are explored for variation as a function of semantic category. We demonstrate that some variation is predictable given the form of *like* a speaker produces.

Keywords: homophones, speech production, frequency effects, representation, homographs

1. INTRODUCTION

Homophones are frequently described as a single word form with multiple distinct meanings [3]. By testing differences in both duration and spectral movement that may vary as a function of which *like* a speaker produces, we contribute to a growing body of literature arguing that homophones, by strict definition, do not exist [8, 9]. Indeed, reproductions of an intended lexical item often pattern similarly in their acoustic realization; however, temporal irregularities on a segmental level suggest that subsequent acoustic realizations of a single spoken word-form are not – or, more accurately, are never – identical.

Gahl [9] argues that some of the variability in production is systematic and comes as a result of lexical frequency. In the same way that relative frequencies can affect the acoustic realization of a given word, members of homophonous pairs (e.g., *time* and *thyme*) are subject to frequency inheritance effects, where lower frequency homophonous items tend to behave more like their higher frequency twins (i.e., shorter in relative duration).

Other work indicates word duration may be further governed by part of speech. Dilts [7] observes that duration of a lexical item may be reliably predicted relative to other items using part-of-speech, when controlling for the number of phonemes that comprise the word-form. His results suggest a continuum wherein different parts of speech are increasingly prone to phonetic reduction. For example, adverbs are more likely to be reduced than both nouns and verbs. While Dilts defines phonetic reduction as any deviation from the canonical form of a word, we extend this definition to suggest that more reduced parts of speech (e.g., adverbs) are then more prone to coarticulation and segmental deletion, further accounting for variation in duration. If so, variation in natural speech would not (entirely) come as a result of physiological and temporal constraints – thus, such a finding suggests factors like part of speech require increased attention in phonetic inquiry.

Table 1: Different semantic/syntactic functions of *like,* with frequency counts (all speakers summed)

FUNCTION	CODE	EXAMPLE	NOTES	FREQ
Verb	V	I like cookies; I would like to go	Any verb	15
Noun	N	to compare like with like	Any nominal use	0
Adjective	Adj	the elf-like maiden	Attributing a quality	0
Adverb	Adv	It tastes like crap	Modifying a verb	2
Quotative	Q	And then he was like "…"	Marks a quotation	139
Discourse Marker	DM	Like, they had scraped her	Place marker, phrase initial	66
Discourse Particle	DP	They had like scraped her	Place marker, phrase internal	299
Approximator Adverb	AA	taken all day to go like 30 miles	Indicates approximate measurement	99
Exemplifier	E	Do you have like a mint or something?	Indicates an example or prototype	177
Conjunction	Conj	He looks like he's seen better days	Connects two clauses	25
Comparative	Comp	He used to have a car like mine	Compares two items	62

One way to disambiguate lexical frequency effects in speech production from those of grapheme frequency and grapheme transparency is to focus on homophones and homographs. Analysing groups of words that share both (broad) phonetic transcription as well as orthographic representation is the most effective way to control for orthographic effects. Drager [8] adopted this approach while exploring the use of *like* from recordings of students in an all girls' high school in New Zealand. For reasons beyond potential frequency effects, *like* provides an ideal focus in this type of study on account of its varied semantic and grammatical functioning. Following previous research on the grammatical uses of *like* [4, 5, 12] the present work identifies 11 sufficiently distinct forms for the investigation of acoustic variance in production due to which function *like* serves at a given time (see Table 1). For example, based on the results in Dilts [7], the adverbial use of *like* would be predicted to be more prone to phonetic reduction than would be the verbal use. Therefore, it seems that if such a predisposition is indeed the case, then adverbial use of *like* should generally be of shorter duration.

We expand upon Drager's work [8] by exploring how lexical and segmental duration, as well as spectral movement within the vowel nucleus (diphthongization), may vary within western Canadian English (WCE) due to *like*'s semantic/ grammatical role in an utterance. Additionally, where in Drager's work [8] it made sense to combine select adverbial forms with the verbal use of *like* as a single lexical category, these forms showed sufficient acoustic differences in our data to be treated as distinct and were therefore analysed separately. Finally, where the population in the previous study [8] was restricted to a single gender, we explore spontaneous speech from both male and female talkers.

2. EXPERIMENT

Data analysed for the present work were collected and extracted from ten conversations of spontaneous western Canadian English.

2.1. Participants

Participants were seven female and three male speakers between 18 and 34 years of age, all of whom were native English speakers from western Canada (i.e., raised in provinces west of Ontario). No known hearing impairments were reported.

2.2. Procedure

Spontaneous conversations were recorded within a sound attenuated booth at the University of Alberta; speakers conversed with unrecorded, off-site interlocutors via a cordless telephone. This method was used to reduce the influence of the laboratory setting and provide a reasonable approximation of everyday, spontaneous speech [17]. Conversations, roughly 25 minutes in length, were captured for each participant using a Countryman E6 head-mounted, condenser microphone; a Korg MR 1000 recorder; and an Alesis Multimix 8 mini mixer for phantom power. All recordings were sampled at 44.1 kHz and 16 bits. Two linguists individually transcribed each conversation; these orthographic transcriptions were then combined into a single master transcription.

2.3. Materials

All instances of *like* were identified and segmented using Praat [2]. The mark-up included both word-level and the segmental boundaries /l + aɪ + kc + k/, where 'kc' represents the velar stop closure and 'k' the subsequent burst release (when present); all items were segmented by a trained phonetician using cues similar to those outlined in Drager [8].

We included 884 productions of *like* in the analysis (ranging from 37 to 181 instances per speaker), however 15 additional tokens were excluded for concurrent laughing, 1 for yawning, and 8 for having no discernible boundary between the burst release and a word-final exhalation.

Acoustic measurements were then extracted for Word Duration, Segment Duration (/l/ + /aɪ/ + /kc/ + /k/), F1 and F2 at both 25% and 75% of the vocalic segment's duration, and speech rate calculated as syllables/second using the target plus adjacent words.

Lexical items immediately preceding and following occurrences of *like* were also coded orthographically, additionally noting their number of syllables; 22 of these adjacent productions were deemed unusable due to laughter, yawning, or repair strategies. If there was an interval of silence exceeding 1 second between the final boundary of a preceding word and the onset of /l/, or the final boundary of /k/ and the onset of a following word, it was presumed to be a speech pause and the adjacent word was excluded from speech rate calculations.

3. RESULTS

The following analyses employed linear mixed effects modelling to explore spectral movement within the diphthong, as well as both word and segment duration. All models were fit using R [14] and the *lme4* package [1]. A backwards stepwise model selection procedure was used with *Subject* as a random-effect predictor and *Age, Gender, Speech Rate*, *Word Duration* (for the segmental analyses), and *Lexical Category* as fixed effects predictors.

3.1 Word Duration

We analysed 884 instances of *like* and the final model included by-Subject random slopes for Speech Rate. Results suggest word duration may reliably be used to predict which *like* a speaker produces. Using the high frequency discourse particles as the model intercept resulted in significantly shorter durations for the approximator adverb (β = -0.1, $t(874)$ = -3.24), comparative adverb (β = -0.11, $t(874)$ = -2.94), conjunctive adverb (β = -0.17, $t(874)$ = -3.29) and quotative (β =

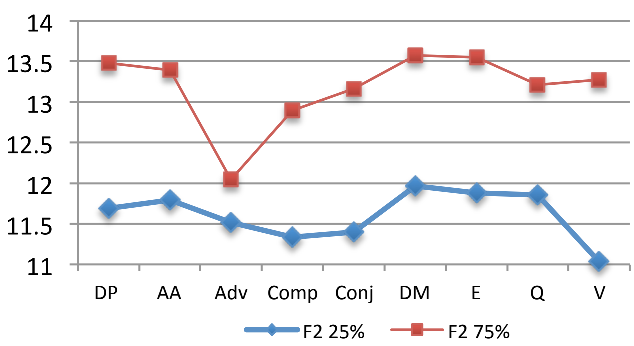
-0.13, $t(874)$ =-4.89). A significant effect was found for Speech Rate (β = -0.05, $t(882)$ = -2.09). We did not find significant effects of Age and Gender.

3.3 Vowel Duration

Vowel duration was a strong predictor of *like*'s function, and the final model also included by-Subject random slopes for Speech Rate. The vocalic segment within the approximator adverb *like* was produced with a significantly shorter duration than the discourse particle (β = -0.05, $t(874)$ = -2.46). Significantly shorter mean durations were also the case for the discourse marker (β = -0.09, $t(874)$ = -3.28) and the quotative *like* (β = -0.05, $t(874)$ = -2.27). Verbal forms of *like* appear to include vocalic segments longer than those of the discourse particle (β = 0.16, $t(874)$ = 2.82). Significant main effects were recognized for both the control variables Speech Rate (β = 0.06, $t(882)$ = 3.07) and Word Duration (β = 0.57, $t(882)$ = 11.18), as was a significant interaction between these two variables (β = 0.04, $t(882)$ = 4.13). We did not find significant effects of Age and Gender.

3.2 Formant Movement

Figure 2: Formant movement at 25% and 75% of F2 in the vocalic segment [aɪ] in BARK



Degree of diphthongization was calculated as Euclidian distance after transforming the formant values into Bark [16]. A difference of zero Bark indicates no movement, whereas higher values (i.e., the summed differences between the movement in both formants of interest) indicate increasing *spectral distance* between the two points of measure. Spectral measurements were recorded at 25% and 75% of the overall vowel duration similar to methods outlined in [13]. Random slopes did not improve the model fit. We found that quotative *like* (Q) was produced with significantly less movement (β = -0.31, $t(874)$ = -5.01) and verbal *like* with significantly more diphthongal movement (β = 0.37, $t(874)$ = 2.29) than the discourse particle (i.e. the intercept). A main effect was observed for Speech Rate (β = 0.16, $t(882)$ = 3.11) and a significant interaction was observed between Speech Rate and (log) Word Duration (β = 0.1, $t(882)$ = 3.61). We did not find significant effects of Age and Gender.

4. DISCUSSION

The present study demonstrates variation in many of *like*'s semantic/grammatical functions. First, there are remarkable differences in the relative frequencies of forms within the dataset – overtly lexical forms are relatively underrepresented. Their scant representation supports *like*'s continued grammaticalization, where the most frequent forms have been semantically 'bleached' when compared to those more lexical in their use.

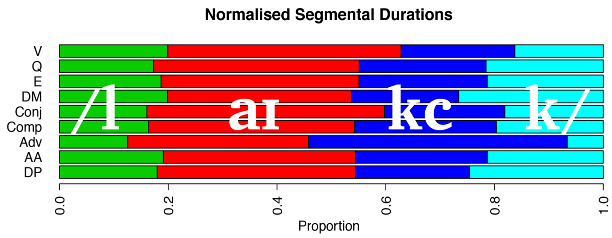
The majority of *like's* forms are based upon an underlying semantic element of comparison. Recognizing verbal *like* as the semantic oddball, inasmuch as this form has little to do with comparison, its low frequency may be explained by way of these semantic underpinnings. It is plausible this growing disuse comes as a result of its semantic disconnect from the other forms. Typical grammaticalization processes predict the verb *to like* may fall out of use in WCE [12, 15].

The grammaticalized discourse particle, however, is considerably more frequent in our dataset than any other form. If frequency effects similar to those described in [9] are the sole driving force behind durational differences, then we should expect this form to be realized with the shortest mean durations, but this seems not to be the case. The discourse particle exists somewhere in *the middle*, where the less frequent approximator adverb, comparative, conjunctive, and quotative forms were all produced with significantly shorter mean durations (recall that based on results in [7] we expect these adverbial forms of *like* to be of shorter duration than verbal *like*). This may be explained by the phrasal-position and grammatical function typically associated with discourse particles. While the approximator, comparative, quotative, and conjunctive forms of *like* are also used phrase-internally, the discourse particle is associated with disfluencies and pauses, and is often used to allow the speaker to plan or inform the listener about turn-taking. It is possible, then, that any such frequency effects may be outweighed by those of phrase-final lengthening.

We have demonstrated durational differences on a segmental level as well. Durations have been normalized in Fig. 2 to reflect mean proportions of each form of *like*, where component segments vary to some degree according to semantic or grammatical function. Secondary sub-phonemic cues, present as durational differences, could be informative to the listener in a manner similar to that of vowel duration and perceived consonant voicing [6]. Just as listeners are more likely to perceive consonants preceded by vowels of longer duration as voiced, it is plausible that listeners could learn to identify and utilize sub-phonemic cues to differentiate homophonous forms. These cues appear to be available for listeners to capitalize on, though it is not yet known to what degree these differences are actually recognized.

Finally, the degree of diphthongisation brought to light important differences between select uses of *like*. Specifically, quotative vocalic segments exhibited less spectral movement than did other forms, and the verbal *like* relatively more movement. One possible explanation for restricted movement in the quotative form lies in phrasal compounding [10], where this function is often 'pre-packaged' with a preceding pronoun and a variant of the verb *to be* (e.g., 'He's *like*...'). It is possible these lexical clusters are recognized and processed by speakers as single units, blurring lexical and syntactic boundaries. If so, these high-frequency collocations/ units would then be produced more quickly and, in turn, be prone to phonetic reduction, as in Lindblom's Target Undershoot model [11] (where temporal and physiological constraints influence a given production). Moreover, this would further harmonize with predictions based on [7] where verbal *like*, with a longer duration, is less prone to phonetic reduction and therefore is allotted more time to fully realize a vocalic segment with more drastic spectral movement – as in the canonical form.

Figure 2: Segmental duration for different forms of 'like' presented as (mean) proportions of the (mean) duration.



5. CONCLUSION

This work contributes to a growing literature that argues homophones are in actuality rather different in production. Whether it be due to effects of lexical frequency, neighbourhood density, an attempt to make distinguishing information available to the listener, or something as yet unexplored, the present work provides support for this argument insofar as the forms of *like* attested in our dataset exhibit systematic, statistically significant differences on both lexical and segmental levels, and in both temporal and spectral dimensions. Though the extent to which listeners are able to utilize such information is as yet unknown, it does appear that speakers seem to distinguish homophones in production.

7. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6.

[2] Boersma, Paul & Weenink, David (2012). Praat: doing phonetics by computer [Computer program]. Version 5.1.43, retrieved from http://www.praat.org/

[3] Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *J. Verbal Learning and Verbal Behavior*, *12*(4), 335-359.

[4] D'Arcy, A. (2006). Lexical replacement and the like (s). *American speech*, *81*(4), 339-357.

[5] D'Arcy, A. (2007). Like and language ideology: Disentangling fact from fiction. *American speech*, *82*(4), 386-419.

[6] Denes, P. (1955). Effect of Duration on the Perception of Voicing. *J. Acoust. Soc. Am*., 27, 761.

[7] Dilts, P. C. (2013). Modelling phonetic reduction in a corpus of spoken English using Random Forests and Mixed-Effects Regression (Thesis).

[8] Drager, K. K. (2011). Sociophonetic variation and the lemma. *J. Phon.*, *39*(4), 694-707.

[9] Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474-496.

[10] Lieber, R. (1992), Deconstructing Morphology, Chicago, University of Chicago Press.

[11] Lindblom, B. (1963). Spectrographic Study of Vowel Reduction. *J. Acoust. Soc. Am*., 35(5), 783-783.

[12] Meehan, T. (1991). It's like, 'What's happening in the evolution of like?': A theory of grammaticalization. *Kansas Working Papers in Linguistics*, 16.

[13] Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *J. Acoust. Soc. Am.*, 80(5), 1297-1308.

[14] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

[15] Traugott, E. C. (1995). The role of the development of discourse markers in a theory of grammaticalization. ichl XII, Manchester, 1-23.

[16] Traunmüller, H. (1997). Perception of speaker sex, age, and vocal effort. *Phonum*, *4*, 183-186.

[17] Warner, N. (2012). Methods for studying spontaneous speech. Chapter in *The Oxford Handbook of Laboratory Phonology*, A. Cohn, C. Fougeron, & M. Huffman (eds.) Oxford Univ. Press. 621-633.