Say What? Why users choose to speak their web queries

Maryam Kamvar, Doug Beeferman

Google Inc, Mountain View, California, USA

mkamvar@google.com, dougb@google.com

Abstract

The context in which a speech-driven application is used (or conversely not used) can be an important signal for recognition engines, and for spoken interface design. Using large-scale logs from a widely deployed spoken system, we analyze on an aggregate level factors that are correlated with a decision to speak a web search query rather than type it. We find the factors most predictive of spoken queries are whether a query is made from an unconventional keyboard, for a search topic relating to a users' location, or for a search topic that can be answered in a "hands-free" fashion. We also find, contrary to our intuition, that longer queries have a higher probability of being typed than shorter queries.

Index Terms: speech, voice, search, Google, query.

1. Introduction

Understanding what factors lead users to speak rather than type input is valuable in the design of applications that offer both input modalities. This insight can guide the user interface designer to make the speech input feature more prominent in certain contexts, activate it automatically in others, or hide it entirely when it is very unlikely to be used. Furthermore, the predictive models used in speech recognition can exploit this type of contextual awareness. For example, the topics and query lengths that are characteristic of speech queries can be used to focus the recognizer's statistical language model in the right places, reducing perplexity and word error rate. To achieve a better understanding of the context in which spoken search applications are used, we present a logs-based comparison of web search patterns across two query input modalities, typing and speaking, in a search environment that offers a straightforward choice between the two.

2. Related Work

Analysis of search patterns in the area of *voice* web search (where the query is entered by speaking rather than by keyboard) is sparse, most likely because spoken web search is a relative newcomer to commercial search interfaces. The three major search engines – Google, Yahoo and Microsoft did not launch voice-enabled web search until 2008[1] [9,10].

Voice search interfaces and architectures have been presented in the past [1,6,8], and the use of voice search logs to improve recognition models has been discussed [2,7]. However, the focus of our paper: the characteristics of speech search queries, and the factors that influence a users' decision to speak a query, has not been discussed extensively at the time of writing.

[1] Google launched Goog-411, a voice search utility for local information, in April 2007, and a multi-modal version of Goog-411 was launched in Google Mobile Maps in July 2008. Microsoft and Yahoo released similar products in 2007, however these products only allowed users to search over a local repository of information, not the entire web.

Figure 1: *The Google Mobile Application*.

3. Google Mobile Application

We analyze search patterns of users who issue both typed and spoken queries from the Google Mobile Application (GMA). GMA is a downloadable search application that also serves as a portal to other Google products. BlackBerry users can choose to enter their query either by typing or speaking. To speak a query they must press the green "call" button on the phone's keyboard, and speak their query while the button is depressed. The release of the green call button signals the user is done speaking the query.

Aside from the query input modality, the "voice" and "typed" search experience is almost identical. The only difference in the search experience is in the "search suggestions". When typing a query, a list of possible completions is provided below the search box. If a user speaks a query, query completions are not provided, but a list of possible alternative recognitions is presented after the query is executed. The search results returned for spoken and typed queries are the same.

4. Dataset

For this study, we considered users on a BlackBerry phone in the United States who issued queries during a 4-week (28-day) period during the summer of 2009. We restricted the users in our data set to those who had issued at least one spoken query, and one typed query. These users amount to over 20% of GMA's US BlackBerry user base. This restriction ensured that our entire user sample was aware of both input modalities.

We randomly sampled approximately 75,000 users (and the over 1,000,000 queries that they issued in both modalities) from the set of users described above. We sampled these users by selecting the log records issued by a random subset of "install ids". An install id is a unique identifier assigned to each GMA client when the application is downloaded, and is

analogous to a browser cookie. All of our data is strictly anonymous; the install ids bear no user-identifying information.

5. Why Users Speak

In this section, we examine if factors such as difficulty of query entry and query topic are predictive of a user's choice to speak the query.

5.1. Difficulty of Query Entry

We examine if the probability a user will speak a query increases as the difficulty of typing a query increases. In this section we consider three proxies for the difficulty of typing a query: keyboard used to type a query, query length, and query popularity.

5.1.1. Keyboard Type

BlackBerry phones have two common keyboard types: a full qwerty keyboard which assigns one letter per key, and a compressed keyboard which assigns two letters for most of the keys. Table 1 shows the percentage of users with each keyboard type.

Compressed keyboards make typed query entry more inefficient because on average more key presses are needed to enter a query. To understand if the type of keyboard available to a user was correlated with a users' decision to speak a query we computed:

```
P(spoken query | keyboard type)
```

As shown in Table 1, there is a much higher probability that a user will speak a query if the keyboard available is "inefficient". A user with a compressed keyboard is 20% more likely to issue a spoken query.

Table 1. *Keyboard type.*

Keyboard Type	% of sampled users	P(spoken query | keyboard type)
Full	86.9	.346
Compressed	13.1	.416

5.1.2. Query Length

In this section we examine another proxy for the difficulty of typing a query: query length. We consider query length to be a proxy for the difficulty of typing a query because longer queries require more key presses. To examine if query length is correlated to a users decision to speak a query, we computed the probability that a query was spoken, given its length. This is expressed as:

```
P(spoken query | query length)
```

If users preferred to speak longer queries (perhaps to realize a time savings in query entry), we would expect this probability to increase as query length increased. On the other hand, if users preferred to type longer queries (perhaps because of the perception that longer spoken queries have a lower probability of being recognized correctly), we would expect this probability to decrease as query length increased.

In Figure 1, we compute the probability a query is spoken as a function of the number of words in a query. The red bars indicate the probability if we only consider spoken queries with recognition confidence score greater than 0.8, and the green dots indicate the probability when we consider all spoken queries. In both cases, there is an inflection point in the probability values when a query is 6 words long. Users seem to apply a non-monotonic cost-benefit analysis in deciding in what mode to enter a query. Users are more likely to speak a query shorter than six words than a longer query. For queries longer than six words, the cost-benefit is inverted and the probability of speaking the queries declines as the query length increases.

One explanation for the falloff in the probability of speaking longer queries is the extent to which users need to remember speech queries in an "articulatory buffer" prior to speaking [10]. According to Sternberg, there is an onset latency (the time before speaking something) that increases with the number of things you have to say, whether measured in words or syllables. Thus, as queries get longer, the onset latency increases, making it less convenient to speak.

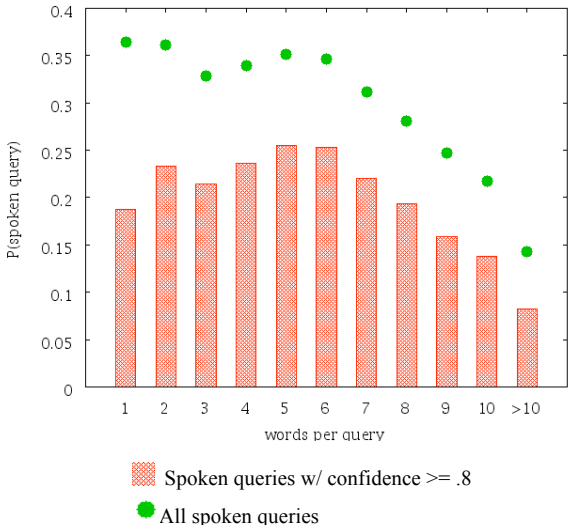

Figure 1: *P(spoken query | query length)*

The inflection point at six words seems to be significant. If we condition these probabilities on the keyboard type, that is if we plot

```
P(spoken query | query length & keyboard
type)
```

for each keyboard type, we find the inflection point remains at six words for both compressed and full keyboards.

A hypothesis that arises from this finding is that queries beyond five words exceed the capacity of the "articulatory buffer" of voice search users. This is in accordance with the Miller's theory that we are limited to processing information in units of "seven, plus of minus two" [11]. Thus, queries longer than five words may exceed a comfortable "articulatory buffer", whereas typed queries of this length can be edited and more easily extended during input. This may be an important indication that the current voice input interface is not suitable for longer input tasks. Interfaces which provide streaming recognition results may be needed before users adopt voice input for longer search queries.

5.1.3. Query Popularity

In this section we continue examining how proxies for query entry difficulty affect the likelihood a user will choose to speak the query. The proxy we investigate in this section is the popularity of a query. Popular queries should be relatively

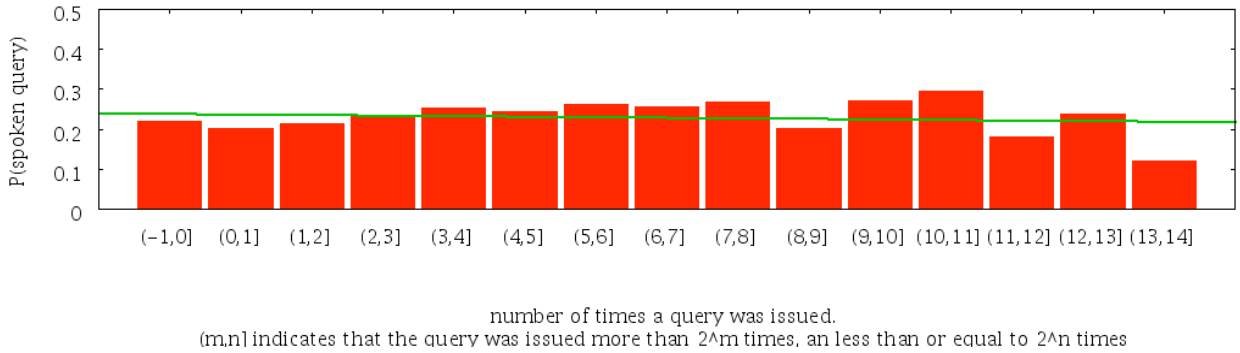

Figure 2: *The probability a query will be spoken for each query frequency, with the best-fit line shown in green.*

easy to type on GMA because of the Suggest feature. The Suggest feature offers likely completions as the user is typing saving users keystrokes and time in entering their query. Thus, in terms of query entry difficulty, we would expect the benefit of speaking an unpopular query to be relatively greater than speaking a popular query.

However, the popularity of a query does not seem to correlate with a users choice to speak it. In Figure 2 we plot, for the spoken queries that have a confidence score of at least 0.8:

`P(spoken query | # times Q was issued)`

The correlation coefficient of the best-fit line (shown in green) is -0.14, indicating no significant correlation between query popularity and probability the query will be spoken. If we consider all queries (not just those with a confidence score of at least 0.8) we still get a very weak correlation coefficient: -0.4.

5.2. Query Topic Classification

In this section, we examine a factor independent of query entry difficulty: query classification. We examine if the type of query issued is correlated with a users decision to speak a query. We classified each query by two different metrics: First, we classified each query into one of 30 different categories. We used the same categorization tool described by Kamvar and Baluja [3], and used in subsequent logs-based studies [4,5]. Next we classified queries by the type of search results returned.

5.2.1. Query Categories

We measured: `P(spoken query | category)` for the 30 different categories described in [3]. The probability a category is spoken is shown in Figure 3.

The queries that have greatest likelihood of being spoken are in the Local, Food & Drink, Shopping and Travel categories. They are all categories that relate to a users' location, or a users' situational context.

Local queries are those whose results have a regional emphasis. They include queries for business listings (e.g. "starbucks holmdel nj") but can also include places (e.g. "lake george") or properties relating to a place ("weather holmdel nj", "best gas prices"). Food & Drink queries are self-descriptive and are often queries for major food chains (e.g. "starbucks"), or genres of food & drink (e.g. "tuna fish", "mexican food"). Both of these query types likely relate to a users' location, even if there is no location specified in the query (this facilitated by the "My Location" feature of GMA which will automatically generate local results for a query based on a user's reported GPS location).

Shopping and Travel queries are likely to relate either to a users' situational context (their primary activity at the time of querying), or to their location. Example Shopping queries include "rapids water park coupons" which may indicate the user is about to enter a water park, and "black converse shoes" which may indicate she would like to compare shoe prices. Queries such as "Costco" and "Walmart" also fall in the Shopping category, but likely relate to a users' location, as the My Location feature automatically generates local results for these queries. Likewise, Travel queries such as "metro north train schedule" and "flight tracker" may relate to a user's situational context, and queries such as "las vegas tourism" may relate to their location.

The categories that are least likely to be spoken are the Adult, Lifestyles and Health categories. Users may choose to type these queries because these categories contain "sensitive" material, which the user would rather keep private. Spoken queries literally "broadcast" a user's search interests.

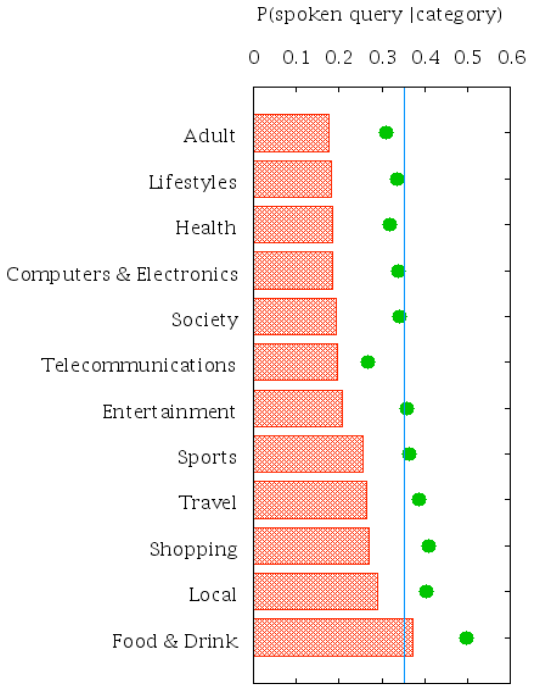

Figure 3: *P(spoken | category). We only show the categories where P(category) > .3%*

5.2.2. *Search Result Categories*

As an alternative to classifying queries themselves, we also classify queries by the types of results that they return. Google's search results may contain "quick results". These results often obviate the need for a user to click on a search result. Sports scores, weather forecasts, and results from specific search verticals (e.g. Image Search, News Search) are all considered "quick results". Although multiple "quick results" may be shown for one query, we classify a query by its top "quick result".

It is interesting to note that on average, spoken queries trigger these "quick results" 12% more often than typed queries. This may indicate that users speak their queries in situations where the entire search experience will be "hands-free". Since queries that present quick results are more likely to contain the users answer on the search page, the user is less likely to need to click on a result, this keeping their search experience "hands-free".

In Figure 4 we examine the probability that a user will speak each "quick result" type. Maps "quick results" are the most likely "quick result" type returned for spoken queries; Approximately 50% of queries that return Maps "quick result" are spoken. While this is significant, it is unsurprising given the query category distribution presented above.

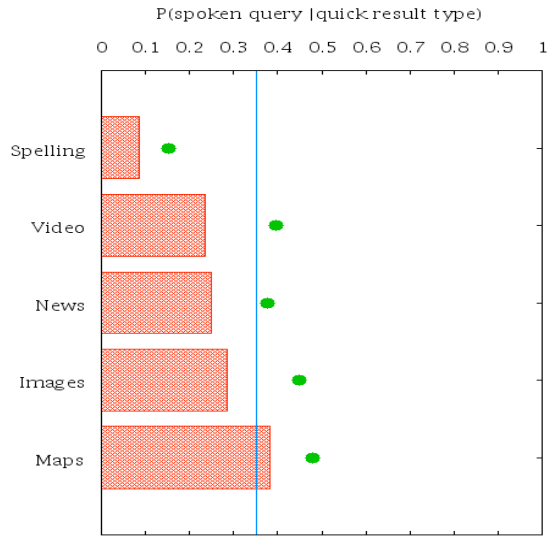

Figure 4: *P(spoken | quick result type). We show the top 5 quick result types triggered*

The next most likely "quick result" type to be triggered by spoken queries is Images. These results often require no further interaction to consume. News and Video, however, most often require an additional click to view the desired information. This finding furthers the hypothesis that users are more likely to choose speech for "hands-free" information needs.

Spelling results (the "did you mean:" result which suggests an alternate spelling for a user's query) are not often triggered by spoken queries. 90% of queries that trigger Spelling results are typed, rather than spoken. This is not surprising since the query recognition backend obviates the need for a user to spell their query, and instead the recognizer is responsible for issuing the well-formed query.

6. Conclusions

The context in which a speech-driven application is used (or conversely not used) can be an important signal for recognition engines, and for spoken interface design. In this study, we analyzed the behavior of mobile search users in an environment that offers a choice between typing and speaking a query. Our goal is to understand the factors that influence a user's choice in query input modality.

We find the factors most predictive of spoken queries are whether a query is made from an unconventional keyboard, for a search topic relating to a users' location, or for a search topic that can be answered in a "hands-free" fashion. We also find, contrary to our intuition, that longer queries have a higher probability of being typed than shorter queries.

Based on our research we offer the following suggestions for improving the voice search experience across mobile devices:

- Improve the handling of local queries, a key factor in the selection of speech as an input mode. Such improvements could take the form of more accurate or more granular location awareness, or better presentation of results for local queries.
- Allow for streaming speech input and real-time recognition results to reduce the cognitive burden on users in order to facilitate longer spoken queries.
- Make it possible to use a larger fraction of result pages in a "hands-free" fashion. When speech is the query input mode, key results (such as the "quick results" in our analysis) can be formatted in a way that reduces the amount of manual interaction required of the user.

7. Acknowledgements

We would like to thank Luca Zanolin, Daryl Pregibon and Johan Schalkwyk for their help and assistance.

8. References

[1] Acero, A. et al. Live Search For Mobile: Web Services By Voice on the Cellphone. N. Bernstein, R. Chambers, Y.C. Ju, X. Li, J. Odell, P. Nguyen, O. Scholz, G. Zweig

[2] Baeza-Yates, R., Dupret, G., & Velasco, J. 2007. A study of mobile search queries in Japan. Query Log Analysis: Social and Technological Challenges. WWW 2007 Workshop.1999.

[3] Kamvar, M., Baluja, S. 2006. A Large Scale Study of Wireless Search Behavior: Google Mobile Search. CHI 2006. 701 – 709.

[4] Kamvar, M. & Baluja, S. 2007. Deciphering Trends in Mobile Search. Computer, 40(8): 58-62.

[5] Kamvar, M., Kellar, M., Patel, R., Xu, Y. 2009. Computers and iPhones and Mobile Phones, oh my! A logs-based comparison of search users on different devices. WWW 2009. 801-810.

[6] Li, X., Nguyen, P., Zweig, G., & Bohus, D. Leveraging Multiple Query Logs to Improve Language Models For Spoken Query Recognition.

[7] Odell, J. & Mukerjee, K., Architecture, user interface, and enabling technology in windows vistas speech systems, *IEEE Trans. Computers*, 56(9):200

[8] http://googleblog.blogspot.com/2008/11/now-you-can-speak-to-google-mobile-app.html

[9] http://blog.vlingo.com/2009/03/about-vlingo.html

[10] Levelt, W. 1898. Speaking: From Intention to Articulation. MIT Press.

[11] Miller, G., 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, vol. 63: 81-97.