Rafe H. Kinsey
Linguistics 110 Paper
Friday, June 3, 2005

Phonological Voice in Whispered Speech

When I first began linguistics and learned phonetics, I would sometimes want to practice sounds in quiet settings, and so I whispered them. I noticed that, even though there was no longer any voicing, I could still distinguish between voiced and unvoiced minimal pairs. In this project, I recorded spoken and whispered speech to examine how the phonological information of voicing is relayed in spoken speech.

I focused on the voiced and unvoiced stops [p] and [b], as an example of a minimal pair in English which contrasts by voicing. I hypothesized that whether or not such sounds were phonologically voiced could be conveyed by information such as stop closure time and aspiration length and the length of preceding vowels, following standard phonological rules for English. Furthermore, from my own observations whispering [p] and [b], I hypothesized that the labial articulation was actually slightly different in the two sounds, and such articulatory difference would be retained in whispered speech.

Finally, recalling Thai's three-way [b], [p], [ph] distinction, I tested a Thai speaker to see whether these sounds were distinguishable in whispered speech. In so doing, I discovered that whispering in Thai is actually very different from English whisper—it is really just quiet spoken speech. An informal survey of speakers of other tone languages revealed that their whispers, too, were just quiet speech.

Method:

I recorded a native English speaker from the Midwest reading a script of sentences, some whispered, some spoken aloud. The recordings were made in the phonetics lab, with the help of Lev Blumenfeld, and were analyzed using the Praat software. The general paradigm was that, for each word, my subject would say, said "Say (test word) now." I tested each of these pairs three times whispered and three times spoken:

- pit, bit
- pat, bat
- bout, pout
- buy, pie
- cap, cab
- cop, cob
- sip, sib

The sentences were ordered randomly, and the participant took short breaks every dozen or so sentences. (I also tested several lesser numbers of examples of other pairs, but will focus on my results for [p] and [b] pairs in this paper.)

Results and Discussion

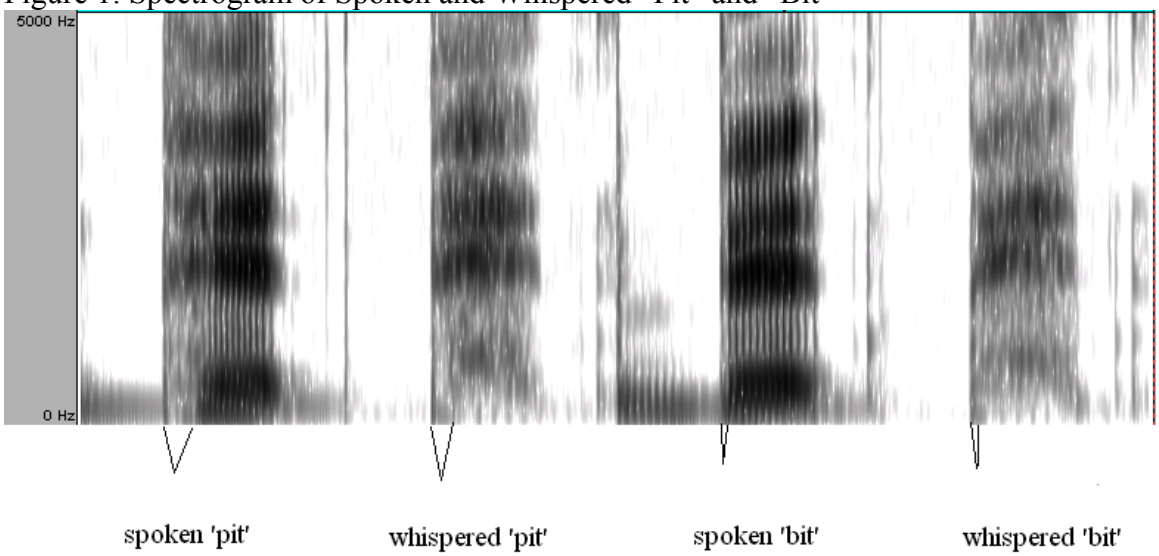
I examined two positions for the [p]-[b] contrast, word-initial (e.g. pit/bit) and word-final (e.g. cap/cab), with several sets of words for each pair. Listening to the whispered recordings, the difference between the b-words and the p-words was usually clear, though much harder to hear than that between the spoken versions.

For word-initial pairs, I found that differences in the length of the closure and in the following aspiration between [b] and [p] were kept to some extent in whispered speech. I discovered in word-final cases that the pattern that preceding vowels are longer when followed by voiced stops was also retained in whispered speech. Finally, I found a clear difference in the formant height of the vowels both preceding and following the stop for [p] and [b], suggesting that they are articulated at slightly different places. This formant difference between [p] and [b] is retained in whispered speech.

Aspiration:

Something corresponding to aspiration can still be seen in whispered [p] and [b]. Consider Figure 1:

Figure 1: Spectrogram of Spoken and Whispered "Pit" and "Bit



Note: The lines point to the area of "aspiration," i.e., the time from release of the lips to the beginning of the vowel.

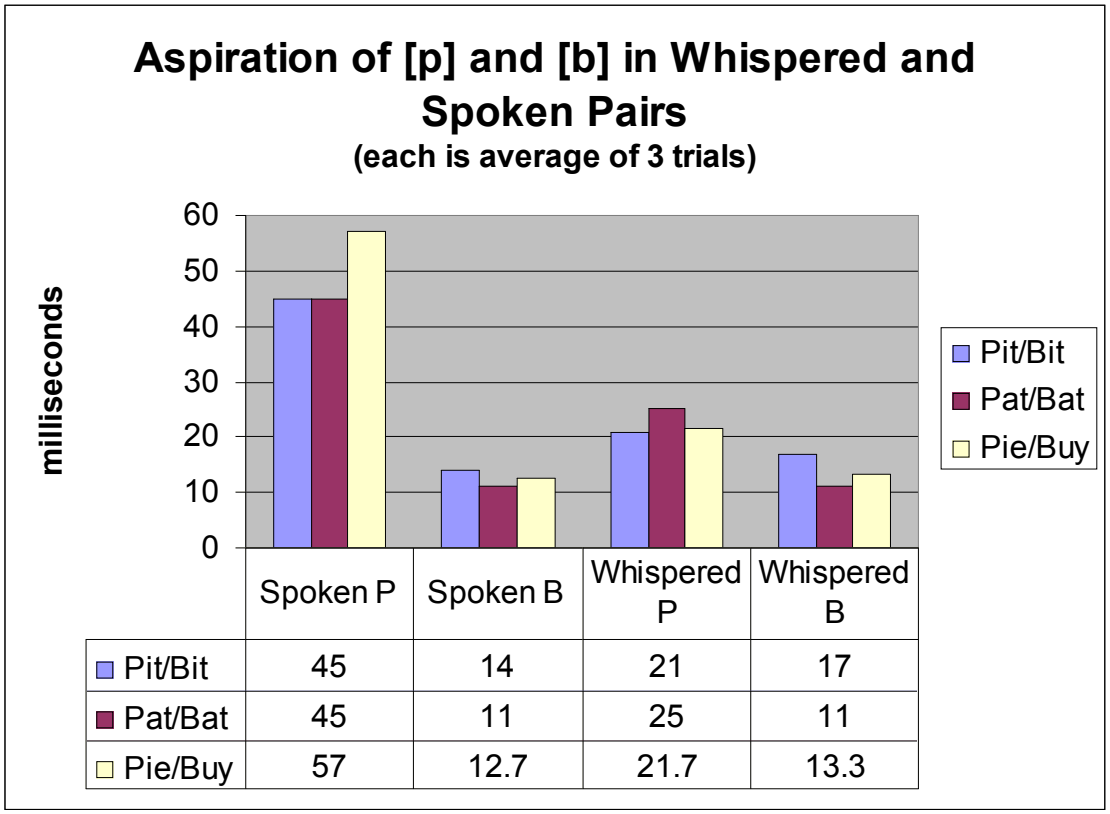
In **pit**[1], there is a clear, long period of aspiration before the vowel begins, which is when the formants become strong. For *pit*, it is harder to tell the difference between aspiration and the ensuing vowel, but there is at least some time, right when the sound

[1] Throughout this paper, I will use **bold** to refer to a spoken word, and *italics* to refer to a whispered word.

begins, where the first formant of the vowel isn't there yet; this can be seen as aspiration. For **bit**, there is a similar but much smaller period of aspiration, which can be seen in the lighter strip before the [i] of the word. Finally, in *bit*, such an area can also be seen, before the first formant solidifies.

Figure 2 shows the data on aspiration for three word-pairs. For all pairs, there is clearly a longer period of aspiration for the whispered [p] than the [b]. However, the similarity between these two, in contrast to the much larger difference between spoken [p] and [b], suggests why the whispered versions sound so much more similar.

Figure 2



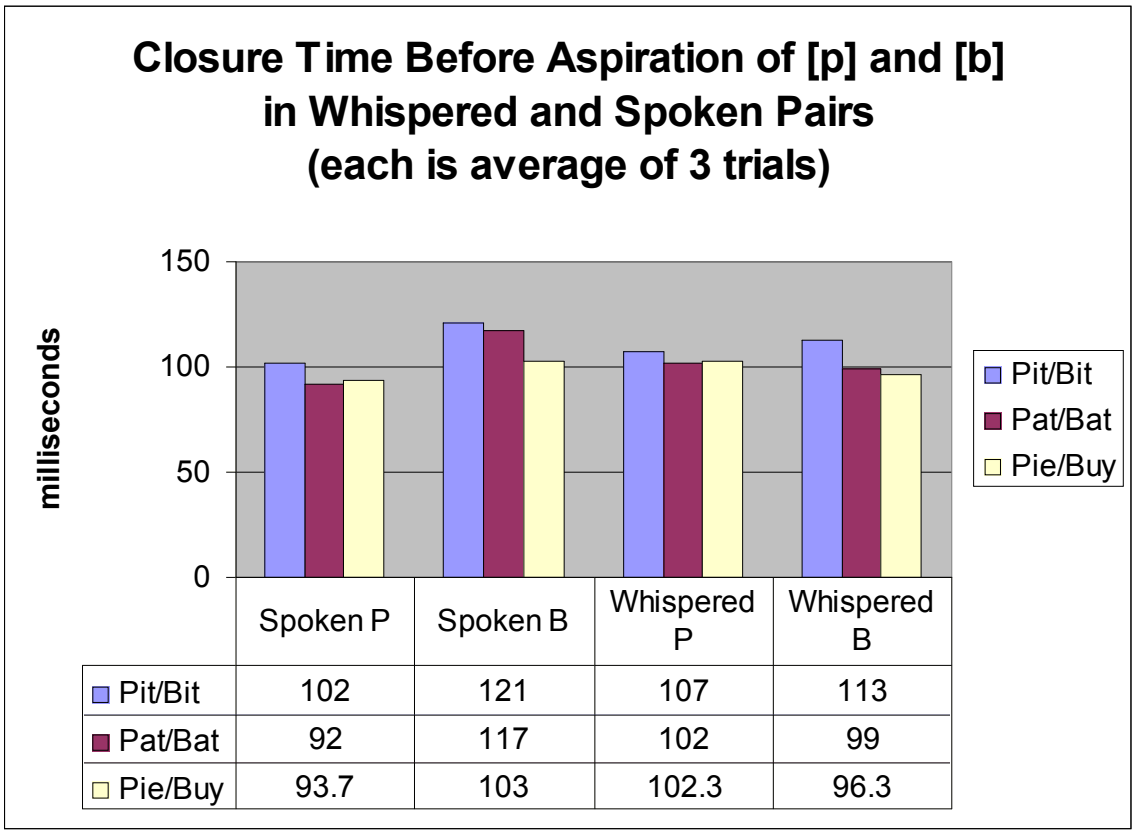
	Spoken P	Spoken B	Whispered P	Whispered B
Pit/Bit	45	14	21	17
Pat/Bat	45	11	25	11
Pie/Buy	57	12.7	21.7	13.3

Note: The aspiration is consistently shorter in [b] than [p]. This difference is carried over in whispered speech, though it is smaller.

There is a significant difference in the time of complete closure of the lips in spoken [p] and [b], with the lips being closed for longer during [b] than [p]. This

difference, unlike aspiration (the time following opening of the lips until the vowel begins), did not remain clearly in the whispered examples (see Figure 3).

Figure 3

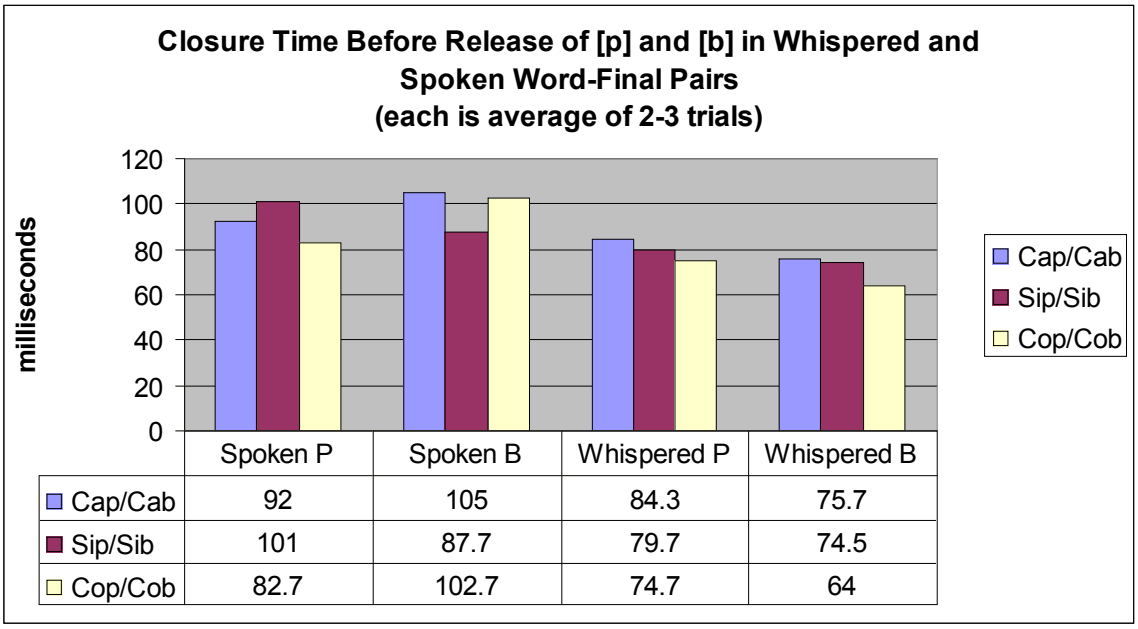


	Spoken P	Spoken B	Whispered P	Whispered B
Pit/Bit	102	121	107	113
Pat/Bat	92	117	102	99
Pie/Buy	93.7	103	102.3	96.3

Note: The difference in length of closure time between [b] and [p] is not consistently retained in whispered speech.

For word-final cases, interestingly, no such clear pattern existed for the spoken words, but one did appear for whispered speech. In whispered speech, the closure time before release for [b] was significantly lower (see Figure 4). Perhaps this is because such a distinction is not made in casual speech because there is more than enough other information for the distinction between the sounds to be made. In whispered speech, though, this basic difference is retained because it is one of the few properties that allows a hearer to distinguish the sounds.

Figure 4



	Spoken P	Spoken B	Whispered P	Whispered B
Cap/Cab	92	105	84.3	75.7
Sip/Sib	101	87.7	79.7	74.5
Cop/Cob	82.7	102.7	74.7	64

Note: Voiced [b] has a consistently shorter closure time in whispered speech, a distinction that does not occur consistently in spoken speech.

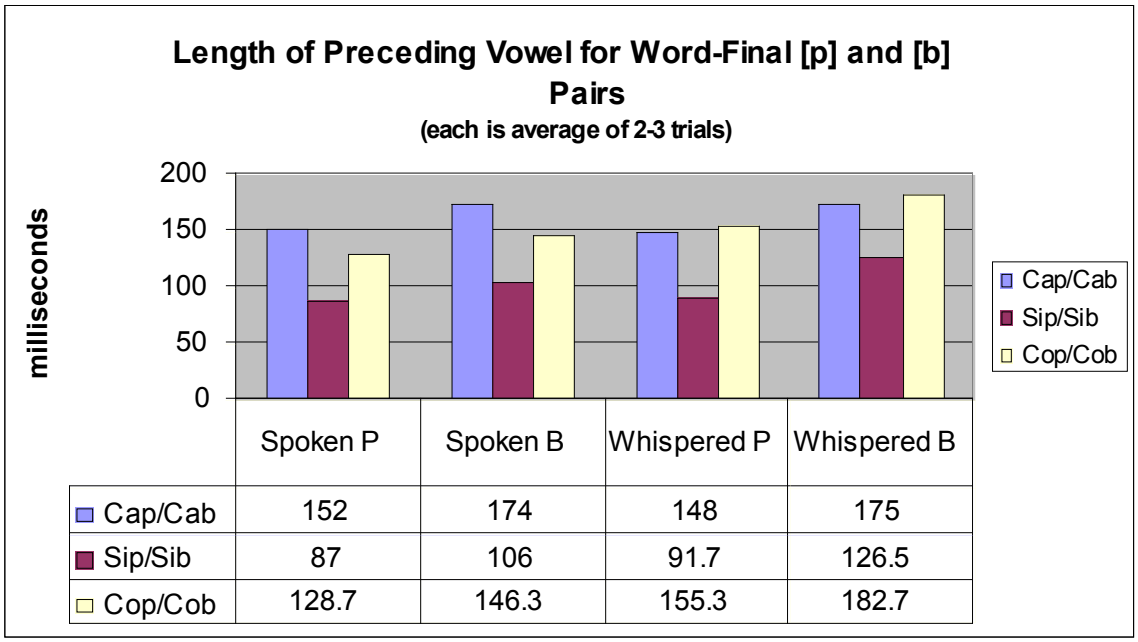
One drawback of the paradigm I used is that the final consonant was followed by the [n] of "now." A better design would have been to follow the word with a vowel (thus the general paradigm being: "Say ____ again," for example), which would have allowed a better analysis of any aspiration for such word-final stops (that is, the time after the [p] or [b] is released but before the vowel begins). In the data I collected, some of the [p]'s and [b]'s were unreleased going into the [n], distorting any such data.

Vowel Length

One regular rule for English vowels is that they are shorter in syllables closed by voiceless consonants than in syllables closed by voiced consonants (Ladefoged 2001: 83). I found this rule continued, very consistently, in the whispered data. Indeed, the difference in vowel length was greater in the whispered pairs (the average difference is 19.5 ms for spoken, and 29.7 ms for whispered). (See Figure 5.) This, again, is evidence

that we emphasize the regular phonological rules in whispered speech to make up for the lack of voicing as a distinguishing feature.

Figure 5



	Spoken P	Spoken B	Whispered P	Whispered B
Cap/Cab	152	174	148	175
Sip/Sib	87	106	91.7	126.5
Cop/Cob	128.7	146.3	155.3	182.7

Note: Vowels preceding voiced [b] are consistently longer than vowels preceding voiceless [p], a distinction that is accentuated in whispered speech.

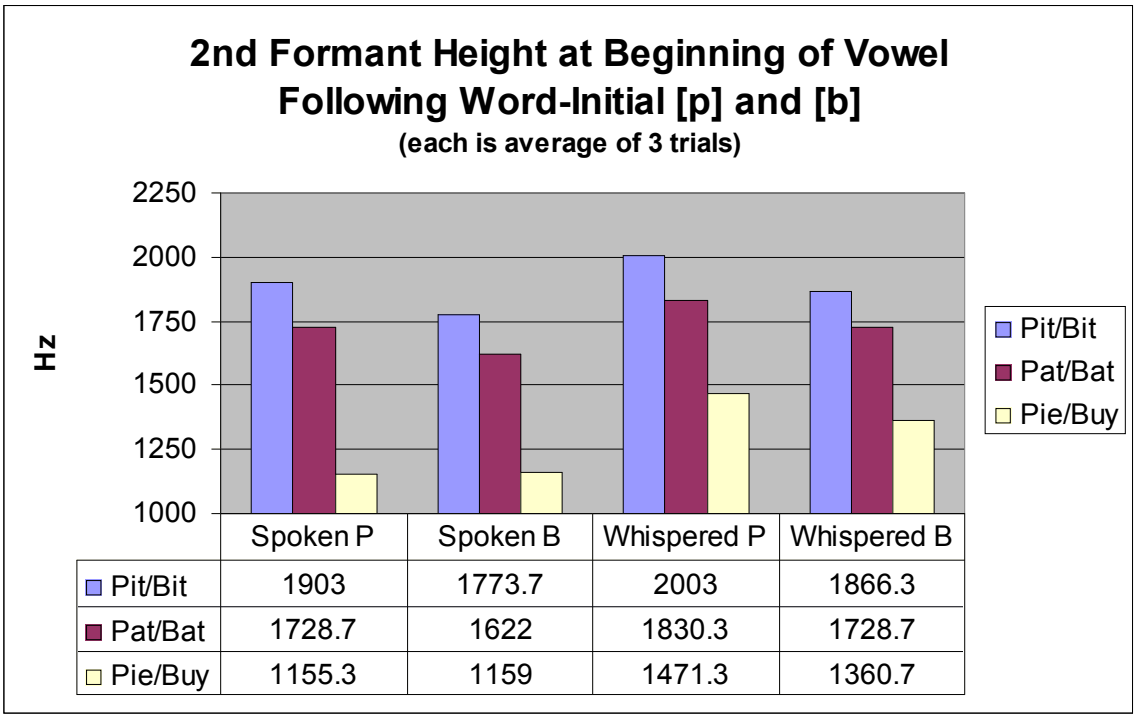
Formant Height

The most interesting result I discovered was that second and third formant heights are consistently lower right before and after [b] than they are for [p], both in spoken and whispered speech. This suggests that there are slightly different lip articulations for [p] and [b], confirming my initial intuitions.

For the word-initial pairs, I took the first clear formant readings made by Praat after the release of the consonant (there were still decently clear formants in the whispered speech, though, as the data show, the formants are consistently a couple of hundred hertz higher in whispered speech; this is presumably due to the acoustic properties of the different type of phonation in whispered speech). For the word-final pairs, I took the last clear formant reading from Praat before the closure. Figures 6

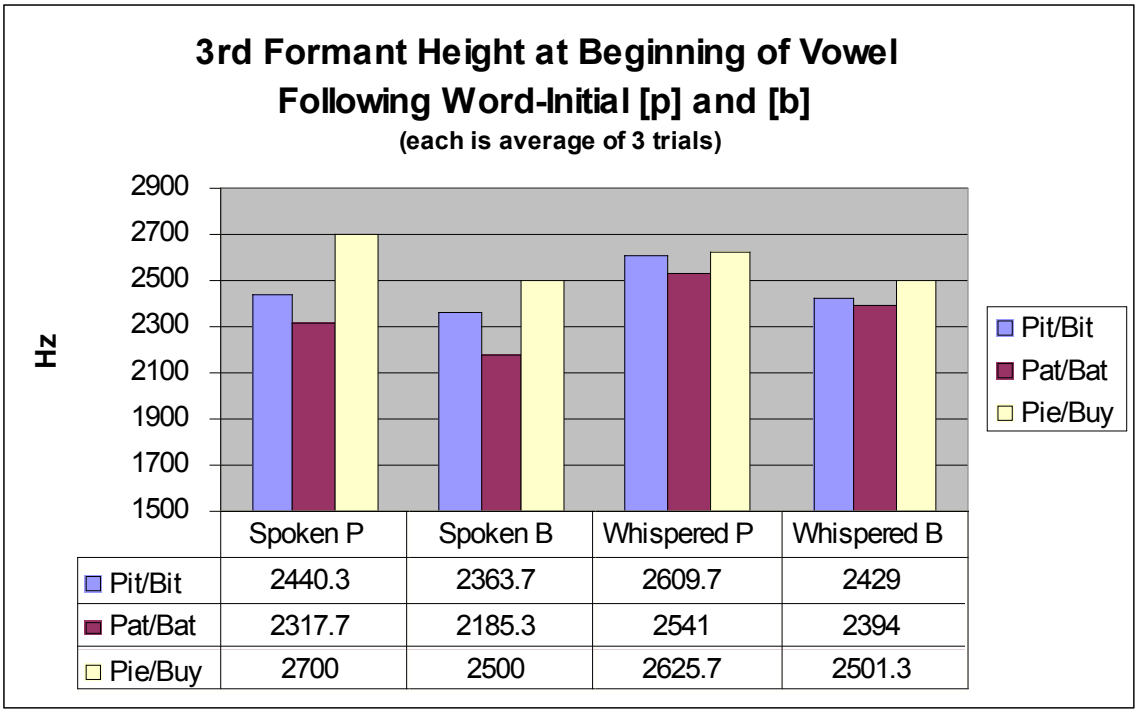
through 9 reveal the clear trend: the formants are consistently lower in the [b] than the [p], for both whispered and spoken speech.

Figure 6



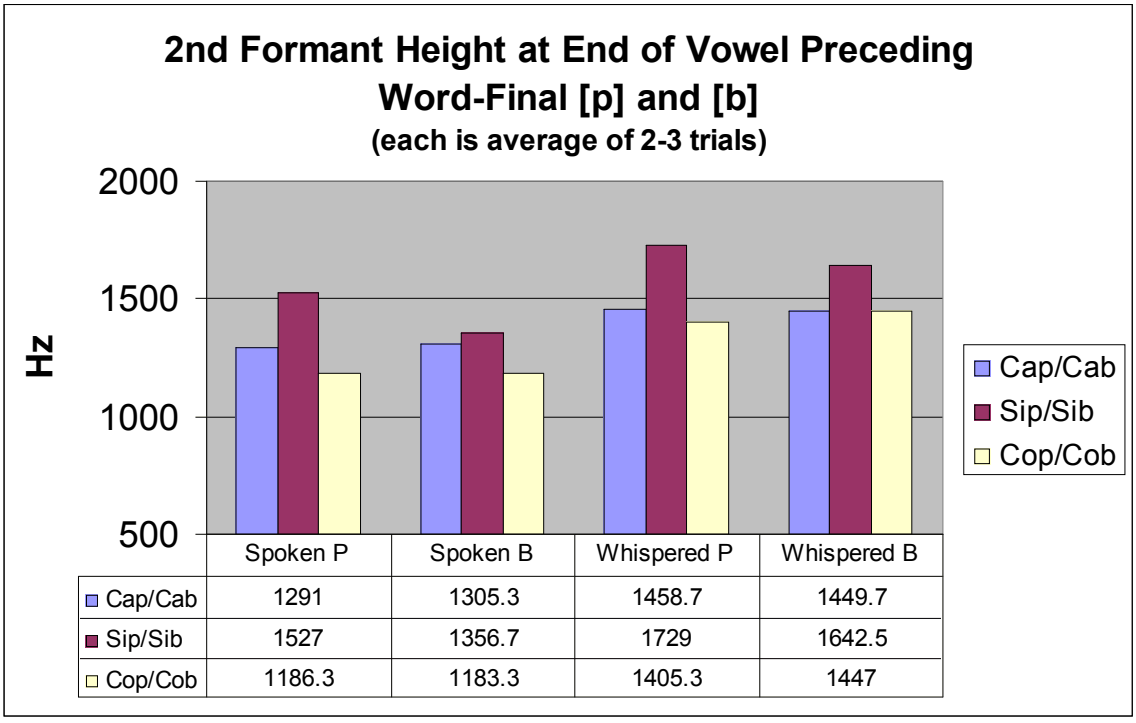
	Spoken P	Spoken B	Whispered P	Whispered B
Pit/Bit	1903	1773.7	2003	1866.3
Pat/Bat	1728.7	1622	1830.3	1728.7
Pie/Buy	1155.3	1159	1471.3	1360.7

Figure 7



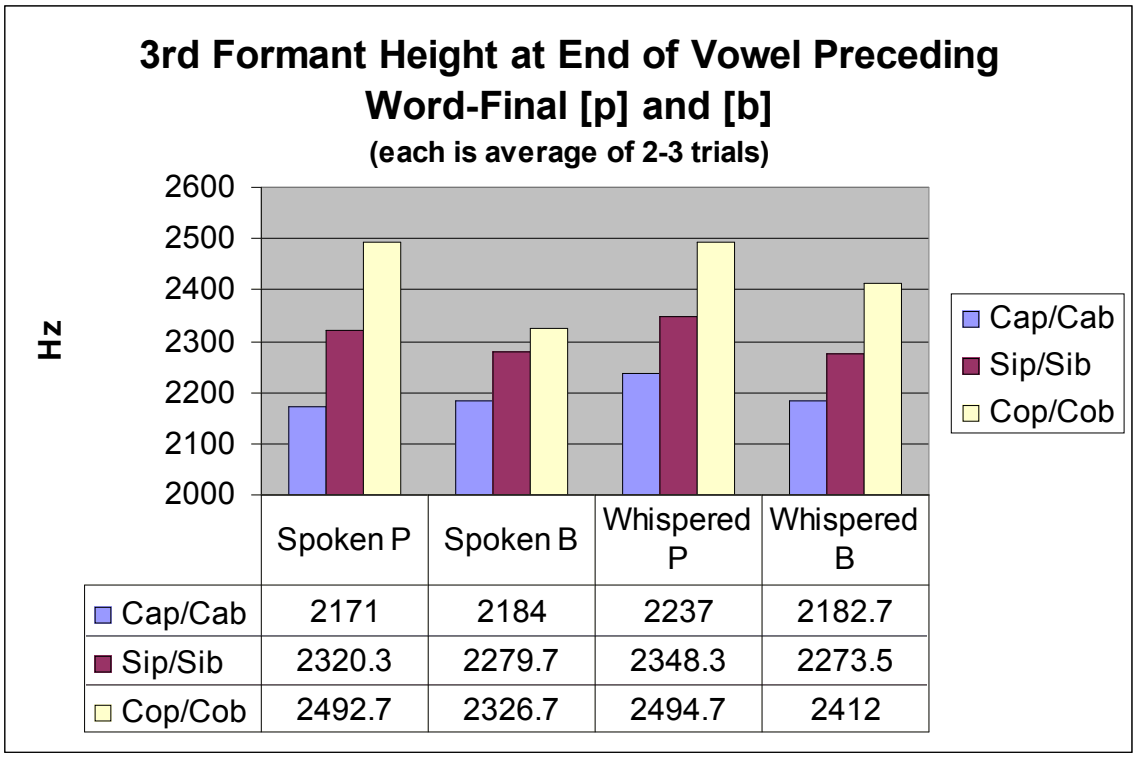
	Spoken P	Spoken B	Whispered P	Whispered B
Pit/Bit	2440.3	2363.7	2609.7	2429
Pat/Bat	2317.7	2185.3	2541	2394
Pie/Buy	2700	2500	2625.7	2501.3

Figure 8



	Spoken P	Spoken B	Whispered P	Whispered B
Cap/Cab	1291	1305.3	1458.7	1449.7
Sip/Sib	1527	1356.7	1729	1642.5
Cop/Cob	1186.3	1183.3	1405.3	1447

Figure 9



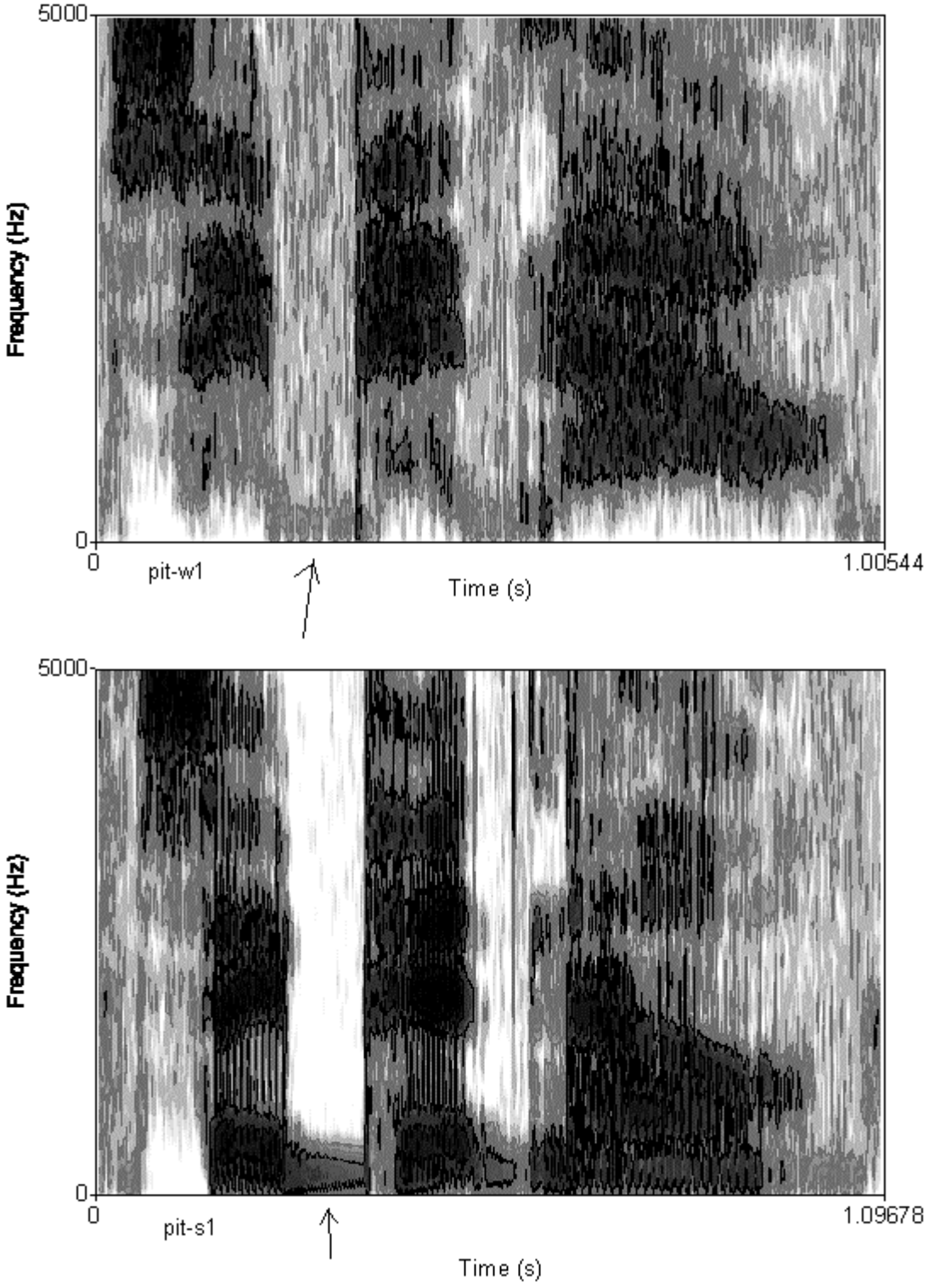
	Spoken P	Spoken B	Whispered P	Whispered B
Cap/Cab	2171	2184	2237	2182.7
Sip/Sib	2320.3	2279.7	2348.3	2273.5
Cop/Cob	2492.7	2326.7	2494.7	2412

For the word-initial pairs, this formant difference is on average slightly larger for whispered speech than spoken speech (Formant 2: 116 ms average whispered difference, 77 ms average spoken difference; Formant 3: 150 ms average whispered difference, 136 ms average spoken difference), continuing in the trend of accentuated differences in whispered speech (which, I would hypothesize, occurs to make up for the loss of voicing information), but this increased difference could also be caused by extraneous factors (such as those factors that make whispered speech have higher formants). Also, there's no such pattern of accentuated difference with the word-final pairs.

Other Explanations

According to Ball and Rahilly (1999: 35) "When whisper is used in this way [i.e., the traditional way to relay a paralinguistic sense of secrecy], speakers transfer voiced sounds to whisper, but maintain voiceless sounds as voiceless (to avoid the loss of contrast between the two groups)." My recordings do not correspond with this view. Consider Figure 10, which shows the spectrograms for "Say pit now" whispered and spoken. The arrows in the two spectrograms point to where the lips have been closed just before the [p] is released. According to Ball and Rahilly's view, there should simply be standard voiceless phonation in both cases (because, they hold, "whispered" voiceless sounds are really just normal voiceless sounds). It is clear, however, that, where the arrows point in the two spectrograms, very different sounds are being produced. In the lower, spoken spectrogram, all the acoustic energy is located below 500 Hz, where a voicing bar would occur; in the upper, whispered spectrogram, there is significant energy much higher, as in a fricative (though not nearly as loud).

Figure 10: Spectrograms of “Say ‘pit’ now” whispered (top) and spoken (bottom).



Note: The arrows point to the closure before the [p] of ‘pit.’ Note the very different amounts of energies, suggesting that these cannot be the same voiceless phonation, as Ball and Rahilly argue.

Other Sounds

I did not collect enough data to examine other minimal pairs, but a cursory analysis revealed similar findings. For instance, among several examples like race/raise and back/bag, the preceding vowel was significantly longer before the voiced sound, in both whispered and spoken speech. Further research might examine whether the formant location differences found in [p] and [b], suggesting slightly different articulations, occur in other sounds.

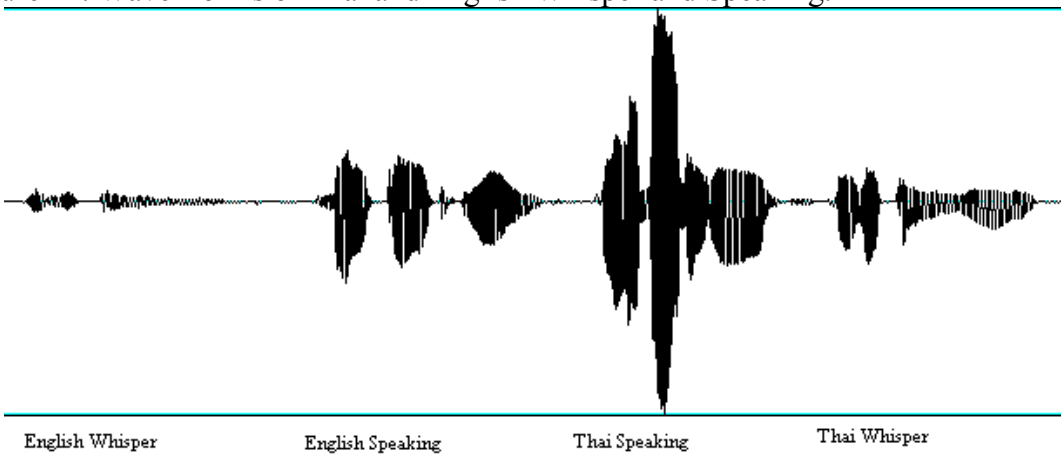
Cross-Linguistic Evidence

Laver (1980: 122) has written that "the use of whisper in a paralinguistic function is very widespread. In English, and perhaps in the vast majority of cultures, to whisper is to signal secrecy or confidentiality."

Recalling the [b]-[p]-[p^h] distinction in Thai, and wondering whether that distinction would be maintained in whispered speech, I brought my roommate, a native Thai speaker, to the phonetics lab and recorded him. He read, either in a whispered or spoken voice, various Thai contrasts, such as [pa]-[ba]-[p^ha], preceded and followed by a wrapper sentence (namely, [p^hudwa ___ dioni], which means "Say ____ now" in Thai). I also tested him on a smaller set of the English "Say ___ now" pairs.

The result was unexpected: his "whisper" for Thai was nothing more than a quiet version of his spoken voice. For the English sentence, however, his "whisper" was like a normal American whisper. Figure 11, which shows the wave-form of whispered and spoken sentences in Thai and English, shows this difference clearly: the whispered Thai looks much closer to the spoken than the whispered English.

Figure 11: Wave-Forms of Thai and English Whisper and Speaking.



Note: It is clear that Thai whisper is far closer to English speaking than to English whisper.

Intrigued, I asked the other two people I knew who natively spoke tone languages, one who spoke Cantonese and one who spoke Mandarin. Both, when I asked them to whisper several sentences in their native tongue, responded in a quiet but definitely not whispered voice, with clear voicing on at least some of the syllables. When asked to whisper in English, however, they both produced normal whispers. Both said this was the standard way to whisper in their native languages. My Cantonese informant couldn't even think of a word in Cantonese for the English "whisper," and found in an English-Cantonese dictionary a Cantonese definition that meant simply "to speak quietly."

Conclusion

In this project, I have examined how the distinction between voiced and unvoiced stops is maintained in whispered speech, focusing on [p] and [b]. I have found three ways that this distinction is relayed: the amount of aspiration and length of closure of the stop; the length of the vowel preceding the stop when the stop closes a syllable; and lower second and third formants for [b] than [p], suggesting a slightly different place of articulation. One interesting pattern I noticed was that some of these differences were accentuated in whispered speech compared to spoken speech. I hypothesize that such

differences are produced in whispered speech because they are necessary to differentiate the sounds, but are less heightened in spoken speech because voicing suffices for differentiating sounds.

The research from this project suggests further research to pursue. Other sound-pairs can be studied (including fricatives); the sounds can be tested in other contexts (e.g., in the middle of words, following and preceding different sounds); and multiple speakers can be examined. The formant differences I noticed in [p] and [b] could be examined more systematically, and non-acoustic methods could study the subtle differences in articulation. Also, further studies of whispering could consider other languages—for example languages with three-way distinctions of [p] [p^h] and [b] like Thai but that are not tonal and thus would have the same whisper as English. Finally, research could examine in more depth the vailidity of Ball and Rahilly's view, which I question, that voiced and non-voiced "whispered" sounds are different phonation types.

Acknowledgments

Thanks to: John Giagnorio and Golf Takapong for letting me record them in the phonetics lab; to Hin Leung and Cheri Li for being informants about Cantonese and Mandarin, respectively; to Shivan Sarin, Alina Lantsberg, Ann Lucena, and Josh Wachspress for humoring my requests to stare at their lips while they repeated words over and over; to David Kinsey for advice on the paper; to Lev Blumenfeld for helping me in the phonetics lab; and to Will Leben and Lis Norcliffe for advice on my project in general.

References

Ball, Martin J. and Joan Rahilly. 1999. *Phonetics: The Science of Speech*. London: Arnold.

Ladefoged, Peter. 2001. *A Course in Phonetics*. 4th ed. Boston: Heinle & Heinle.

Laver, John. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge Unviersity Press.