UCLA Working Papers in Phonetics, No. 108, pp. 164-187

Comparison of speaking fundamental frequency in English and Mandarin

Patricia Keating and Grace Kuo

(keating@humnet.ucla.edu,gracekuo@humnet.ucla.edu)

ABSTRACT

To determine if the speaking fundamental frequency (F0) profiles of English and Mandarin differ, a variety of voice samples from male and female speakers were compared. Two methods of pitchtracking were also compared. Differences in the F0 data from the two methods were small, except that the lowest creaky F0 values were better detected manually. The F0 profiles of English vs. Mandarin speakers were sometimes found to differ, but these differences depended on the particular speech samples being compared. Most notably, the physiological F0 ranges of the speakers, determined from tone sweeps, did not differ between the two languages, indicating that the English and Mandarin speakers' voices are comparable. Their use of voice pitch in single-word utterances was, however, quite different, with the Mandarin speakers having higher maximums and means, and larger ranges, even when only the Mandarin high falling tone was compared with English. In contrast, for a prose passage, the two languages were more similar, differing only in the mean F0, Mandarin again being higher. The study thus contributes to the growing literature showing that languages can differ in their F0 profile, but highlights the fact that the choice of speech materials to compare can be critical.

I. INTRODUCTION

Differences in the characteristic ranges and levels of speaking fundamental frequency (F0) occur within social groups, often tied to cultural-specific notions of politeness (Crystal, 1969; Graddol and Swann, 1989). As a result, differences between male and female speakers within a language can go beyond what physiology alone would suggest (Graddol and Swann, 1983; Henton, 1989). That is, speaking F0 is to some extent an arbitrary aspect of speech. Not surprisingly, then, several studies over the past decades have begun to establish that speakers of different languages or dialects may likewise use characteristically different ranges and typical values of speaking F0 (see Dolson (1994) for a review). That is, a particular pitch range may be part of the phonetic structure of a language, such that, in the limit, a speaker would sound non-native (have a foreign accent) if speaking in a different pitch range.

Perhaps the earliest experimental test of language differences in speaking F0 was a pair of studies in the 1960s. Hanley et al. (1966) compared the medians (rather than the means; see Baken et al. 2000 for comparison of these measures) and standard deviations of the speaking F0s of male UCLA students who were native speakers of English, Spanish, or Japanese reading the Rainbow Passage (from Fairbanks, 1940) and speaking spontaneously for one minute. For both speech samples, there was a main effect of language on the median pitches, but not on the standard deviations. While no post-hoc tests were reported to clarify which language differences

were significant, the medians from the Rainbow Passage were highest in Japanese and lowest in English, while those from the unscripted speech were highest in Spanish and lowest in English. Hanley and Snidecor (1967) followed up with a similar study of female English, Spanish, Japanese, and Tagalog speakers reading the Rainbow Passage, but in this case there were no significant language differences, though there was a trend for English and Tagalog to have lower median F0 than Spanish and Japanese. In sum, the results were mixed, with the only clear result that the English males had the lowest median F0. However, the values reported for English were unusually low in comparison with other studies of English summarized in Baken and Orlikoff (2000).

A number of later studies have compared just Japanese and English speakers (Loveday, 1981; Yamazawa and Hollien, 1992; Ohara, 1992; Todaka, 1993). Other studies have compared other languages: Polish vs. English, Majewski et al. (1972); British English vs. German, Mennen et al. (2007, 2008); Mandarin vs. Min (Taiwanese), S. Chen (2005). Chen compared speaking F0 range (95^{th} percentile), mean, and standard deviation from prose passage reading by Taiwanese Mandarin vs. Min speakers, but found no F0 differences between these two groups of speakers. She then suggested that both Mandarin and Min showed wider F0 ranges than seen in most previous studies of English (though not in all); in addition, inspection of her tables on p. 3228 indicates that the mean F0s are at the low end of published values for English.

Even different dialects of the same language can differ. Deutsch et al. (2009) compared female speakers of Standard Mandarin from two different villages, and found that the "overall pitch levels" differed by about 30 Hz. Comparison of Taiwan vs. Beijing Mandarin speakers (Torgerson 2005) found that Taiwan Mandarin maximum F0 and median F0 were significantly lower, but not minimum F0 or F0 range (though he excluded the very high and very low measurements), and that these differences held for every tone. The F0 range for the falling tone (Tone 4) was especially large in Beijing Mandarin.

Other studies have directly compared Mandarin and English. G. T. Chen (1972) compared the mean, standard deviation, and range of F0 of 4 English and 4 Mandarin speakers (2 males and 2 females each, i.e. a very small sample) reading words and sentences. The Mandarin speakers, especially the women, had wider F0 ranges and larger standard deviations; the Mandarin women's means were lower, while the men's were the same as the English. These results are broadly in line with S. Chen's. In contrast, Eady (1982) compared several measures of F0 from passages read by male Taiwan Mandarin and English speakers. The mean F0 and measures of F0 fluctuation (dynamic movement) were all greater in Mandarin, but the standard deviation (taken as the measure of F0 range) was the same in the two languages. Xue et al. (2002) compared the F0 mean, standard deviation, minimum, maximum, and range of younger and older bilingual speakers (both male and female, but analyzed together), and found that while the older bilinguals had no differences between their Mandarin and English, the younger bilinguals had lower minimum F0 and larger F0 range in their Mandarin (with no differences in maximum F0, mean F0, or standard deviation). Finally, Mang (2001) compared the means for speaking and singing longitudinally for 8 pre-school girls who were either monolingual English, or bilingual (English-Mandarin or English-Cantonese). The speaking F0 decreased over time for both language groups, but was lower in Chinese than in English from ages 2 to 5. However, when the girls were 5-6 years old, the English speaking F0 dropped below the Chinese. In sum, there is some evidence that in Mandarin the F0 range is greater, but results about the mean F0 are mixed.

Of these studies, those which compared bilingual speakers in both their languages (e.g. Yamazawa and Hollien (1992), Ohara (1992), Todaka (1993), and Xue et al. (2002)) are especially valuable, as they clarify that differences between languages need not be due to physiological differences between monolingual speakers of those languages. Even if the differences originated physiologically in the past of each speech community, individual language learners can learn that certain pitch settings are appropriate for one language as opposed to another.

Another idea about language differences is that they may originate in phonological inventory differences. For example, if one language has more voiceless obstruents, or more high vowels, than another, and if voiceless obstruents and high vowels have a raising effect on pitch (Lehiste 1970), then those two languages might well have characteristic pitch differences, though they would likely be small. Yet another proposal (e.g. S. Chen, 2005) is that tone languages will have larger pitch ranges and/or higher average pitches than non-tone languages, on the assumption that lexical tones require a greater extent of pitch than intonation alone does. Chen's discussion suggests that the locus of such a tone-language effect could lie specifically with the occurrence of a high level tone, as in Mandarin and Min. The typical pitch of a high level lexical tone might be systematically higher than that of high intonational tones in a language like English, or there might simply be more lexical high tones in running speech in a tone language than there are intonational high tones in running speech in a non-tone language. Liu (2002) looked at the F0 range in Mandarin by tone, and found that from Tone 1 through Tone 4, the tones have progressively larger F0 ranges. Thus an alternative scenario is that the more Tone 4s in Mandarin speech, the more likely it is to show a large F0 range.

However, the hypothesis that tone languages as such have an *overall larger pitch range* is not supported by Eady (1982), who found no difference between English and Mandarin F0 standard deviations (the measure of range in that study). Similarly, although Chen (2005) interpreted Japanese's lexical pitch-accent as making Japanese somewhat like a tone language, there was no difference between these languages for the standard deviations in Hanley et al. (1966) or Hanley and Snidecor (1967).

The hypothesis that tone languages have an *overall higher average pitch* receives only limited support from the studies reviewed above. Eady (1982) and (for one age group) Mang (2001) found that the mean F0 was higher in Mandarin than in English, but other studies comparing Mandarin and English gave different results. And, if we follow Chen (2005) in considering Japanese to be somewhat like a tone language, explaining its higher median F0 in the Hanley et al. (1966) study, we must note that even in Hanley et al. (1966) it is not clear that the Japanese F0 was higher than the Spanish (Spanish being a non-tone language like English), and the Hanley and Snidecor (1967) follow-up study with female speakers found no overall language differences.

The present study examines the idea that Mandarin and English have different characteristic pitch properties, and that these differences are related to the tonal nature of Mandarin. In addition, the type of speech sample is varied, so that any language differences can be understood in a broader context.

The research questions of the present study were the following: (1) Do English and Mandarin F0 profiles differ? (2) If so, how does that difference relate to the fact that Mandarin is a lexical tone language? In addition, there were two methodological questions: (1) How does the type of voice sample elicited affect the F0 characterization? and (2) How does the method of F0

estimation affect the F0 characterization? We compare a variety of voice samples from male and female speakers of English and Mandarin, calculating F0 in two ways.

II. METHODS

A. Speech samples

Previous literature distinguishes between a speaker's physiological F0 range – the maximum F0 range the speaker's voice is capable of producing – and speaking F0, or the range of F0s a speaker habitually produces in normal speech. Obviously the former is much larger than the latter. We followed previous practice in recording speech samples for estimating both of these kinds of F0 range. For estimating speaking F0, we included more than one kind of speech sample: isolated words, connected read speech in a neutral style, and connected read speech in a lively style. Baken and Orlikoff (2000) review the literature comparing read with spontaneous speech and conclude that because the mean F0 is only slightly higher in reading (about .5 to 2 ST), with no large differences in variability, F0 measures from read speech are representative of more natural speech. More recently, Torgerson (2005) found no differences between read and spontaneous Mandarin speech, reinforcing Baken and Orlikoff's conclusion.

1. Maximum F0 ranges: tone sweeps

a. Unprompted sweeps. The first productions collected were for maximum (physiological) F0 ranges. Two sets of tone sweeps were produced by the speakers. First, speakers produced sweeps following the general procedure described by Honoroff and Whalen (2005). In their experiment, speakers produced an [a] beginning at a "comfortable (self-determined) habitual pitch", and then "increasing fundamental frequency continuously until modal phonation could no longer be sustained". Then, again beginning at a comfortable pitch on [a], they decreased F0 "until the speaker was no longer able to sustain phonation at the low end of the range". In our experiment, English-speaking participants were instructed somewhat differently, as follows (and similarly for the Mandarin speakers, but translated into Mandarin):

- You will record a series of "ah" sounds in which you let your voice sweep over a wide range of pitches (tones). The goal is to see how wide a range of pitches your voice can ***comfortably*** cover, without straining.
- Start by taking a big breath
- Say "ah" at a comfortable, normal pitch
- Then move gradually (but quickly) higher until you feel your voice break. Going into "falsetto" is fine.

They then listened to a demo recording of a rising sweep by the first author. After some practice, three rising sweeps were recorded. Falling sweeps were instructed, practiced, and recorded in the same way. Specifically, speakers were instructed to "move gradually (but quickly) lower until you feel your voice break or give out". In the recorded demo that subjects heard, there was no creaky voice, with the lowest pitches instead very breathy. In the literature, F0 elicitations and/or measurements generally exclude low-pitched glottal creak or creaky voice. However, fluent speech often includes creak. Therefore we elicited a third, new, kind of F0 sweep, which explicitly asked speakers to produce low F0s in creaky voice. Speakers were asked to "move

gradually (but quickly) lower, letting your voice "creak", until you can't go any lower". The recorded demo was very creaky. This third type will be called creaky sweeps. We refer to all three of these types of sweeps as unprompted, because each speaker determined how to perform his or her own sweeps.

b. Prompted sweeps. Based on the discussion in Baken and Orlikoff (2000), the fast-glissando procedure described in Reich et al. (1990) was also used to elicit speakers' F0 ranges. Reich et al. found that this method resulted in larger ranges than other methods that they compared. In our experiment participants heard, and simultaneously imitated, two rapid tone glides. The glides were 4-sec sawtooth waves whose F0s varied linearly as in Table I, with different ranges for men and women. The way in which speakers imitated the tone glides was as follows. They were instructed: "you are asked to imitate those sweeps as much as possible. It's not expected that any single voice can cover the whole range, but do the best you can. You should imitate the tones *while listening to them*." They then clicked on an icon for a tone glide, and practiced imitating it. They recorded three rising sweeps followed by three falling sweeps. We refer to these two types of sweeps as prompted, meaning that each speaker heard and imitated a tone glide for each production.

TABLE I. Tone glide F0 ranges (beginning and end values), taken from Reich et al. (1990).

	rising	falling
men	277-1108	277-69
women	392-1568	392-98

2. Speaking F0 ranges

Three sets of materials were used to determine the speaking F0 range and other properties of an F0 profile in real speech, specifically in reading.

a. Isolated words. A monosyllabic word was produced in isolation. For English, there was one word, *sure* (/ʃɝ/). It was produced in 4 different ways, chosen to produce different pitch contours and ranges (and yet be understandable to naive readers), and repeated 3 times each way. Two of these will be analyzed here. They were described, and the prompts were written, as follows:

1. Normal pitch: In a regular way, at a normal comfortable pitch: —Sure.
2. Excited exclamation: An exclamation with more and more excitement, and higher and higher pitched voice: —Sure! —Sure!! —SURE!!!

For Mandarin, 4 different words were used, one for each of the 4 lexical tones, which involve different pitch contours. They were:

- High level 師 (/ʂl ˥/, *teacher*)
- Mid rising 十 (/ʂl ˧˥/, *ten*)
- Low falling 使 (/ʂl ˩˧/, *to make (someone do something)*)
- High falling 示 (/ʂl ˥˩/, *to show*)

These will be referred to as “shi” words. They were produced in three different ways, two of which will be analyzed here:

1. Normal pitch: same as for English
2. Exclamation: With a higher tone, as if there is an exclamation mark after the word:

師! 十! 使! 示!

Note that, while an Exclamation condition was elicited in both English and Mandarin, only the English version encouraged an increase in excitement. Therefore only the first token from the exclamations will be analyzed here, to ensure greatest comparability between the languages.

b. Rainbow passage. Following many previous studies, a reading of the Rainbow passage (from Fairbanks, 1960) was also obtained from each speaker. The original English text was translated into Mandarin for the Chinese speakers. Speakers read the text through silently before recording. The English version includes 330 words in 19 sentences, and takes about 2 minutes to read. The Mandarin version includes 444 characters in 19 sentences. Of these, 65 are Tone 1 and 131 are Tone 4. Speakers first read this passage silently to familiarize themselves with it before reading it aloud for recording.

c. Little Red Riding Hood story. Third, a livelier prose passage was recorded. The story of Little Red Riding Hood was chosen because it contains dialog from a variety of characters, and is familiar to both English and Mandarin speakers from childhood. A shortened version of the story, which nonetheless preserved a wide variety of character dialog, was produced by consulting the Little Red Riding Hood Project website (http://www.usm.edu/english/fairytales/lrrh/lrrhhome.htm). Our English version of the story (see Appendix) has 857 words in 54 sentences, and takes about 4-5 minutes to read. The Mandarin version has 1086 characters in 68 sentences. Speakers read this story twice, the first time in a neutral voice without acting out the characters (this served as familiarization), and the second time more dramatically, “as you would read it to a small child, in a story-teller voice, acting out the dialogs with different voices for the different characters”. The dialog was printed in a different color for each character. A small subset of the second reading of the story, focused on characters’ dialog, was analyzed for the current study, indicated by boldface in the Appendix. In the Mandarin version, this selection contained 218 tones, of which 29 were Tone 1, and 57 were Tone 4.

B. Speakers

23 American English and 23 Mandarin speakers were recorded, with the goal of having 20 of each language after recording problems, reading errors, etc.. Most were UCLA students in their late teens or early twenties, though a few were older staff or visitors. The Mandarin speakers were self-described as native speakers, and their recorded speech was later verified as sounding native by the second author. All but 4 of the Mandarin speakers were from Taiwan; of the 4 mainland speakers, two were men and two were women. Speakers responded to a solicitation to “Read a few words and a short text; read the story of Little Red Riding Hood using different

voices for the different characters; make high and low voice pitches", and they were paid for their time.

C. Recording procedure

In a soundbooth in the UCLA Phonetics Laboratory, participants were seated individually in front of a laptop computer which presented the instructions for the recording session as a Powerpoint slideshow. They wore a Shure head-mounted microphone, and for part of the experiment, an earbud in one ear. The microphone signal was recorded direct-to-disk on another computer located outside the soundbooth, at a 44.1 kHz sampling rate and a 32 bit quantization rate, using an AudioBox and PCQuirerX. An assistant outside the booth operated the recording computer during the session. Participants paged through the Powerpoint slideshow, practicing each type of production as needed, and indicated to the assistant when they were ready to record each one. The assistant then saved each recording as a .wav file. The assistant did not provide feedback or correction during the recording session. Each recording session took about an hour.

The order of materials in the recording session was the same for every speaker, viz.: unprompted rising sweeps, unprompted falling sweeps, unprompted creaky sweeps; prompted rising sweeps, prompted falling sweeps; isolated words; Rainbow Passage; Little Red Riding Hood in neutral voice, Little Red Riding Hood in story-teller voice.

D. F0 analysis

1. Pitchtracking

Fundamental frequencies of the tone sweeps and isolated words were measured using one or both of two different programs and methods. Initially both were used on a given corpus, so that they could be compared.

The first method used the cepstral pitchtracker in the PCQuirer/Pitchworks program, with manual adjustments of parameters to give the best possible pitchtrack for each utterance. For the tone sweeps, which often covered a wide range of F0s, adjustments were sometimes made for sub-sections of a sweep, so that both lower and higher pitches could be accurately tracked. Pitchtracks at 5 msec intervals were logged to a text file (and if done separately for subsets of a file, later combined into a single pitchtrack for the file). Then, in Excel, the minimum and maximum F0 were found for each pitchtrack. These values were hand-checked in two ways. First, high F0 values were checked to be sure that they did not come from signal artifacts or from fricative segments. Second, waveforms were compared with pitchtracks to identify any creaky intervals where the pitchtrack had failed. For these intervals, the minimum F0 was calculated directly from the waveform (f=1/T).

This first method thus used a substantial amount of hand-processing of files to yield the best possible set of F0 values. Nonetheless, this pitchtracker returns a zero value when no F0 can be found, and many stretches of voiced speech remained untracked, even after adjustments of tracking parameters. Because of these missing values, mean F0 or its standard deviation could not be reliably obtained. We will refer to this method as "cepstral plus manual".

The second method used the STRAIGHT algorithm (Kawahara et al.1998) incorporated into VoiceSauce (Shue, Keating, and Vicenik, 2009), a new program for voice analysis. Other than setting the maximum expected F0 for a folder of utterances to a reasonable value, no

parameter adjustments were made and the program ran entirely automatically; however, some pre- and/or post-processing was required. Specifically, first the "To PointProcess (periodic, cc)" function in Praat (Boersma, 2001) was used in a script to identify target voiced portions of files (tone sweeps, vowels in isolated words, all voiced intervals in read passages) and segment them in a Praat TextGrid file. Another Praat script then displayed each utterance and its TextGrid for manual checking. At this stage, utterances with recording artifacts, and errors in Mandarin tone production, were removed from further analysis. The audio and TextGrid files were then input to VoiceSauce for acoustic analysis. VoiceSauce computes many measures, but here only the STRAIGHT F0 will be reported. STRAIGHT pitchtracks were output either as text, or in the format for an Emu database (Harrington, 2010).

Emu databases were made for the sweep and word corpora, with F0 values at 1 ms intervals. For each Emu database, the audio files were displayed along with their labels and the F0 track, and examined for gross errors of pitchtracking. The original Praat segmentation/ labeling was manually adjusted at this point to modify, or eliminate from further analysis, segments with pitchtracking problems. In general, segment labels were removed from segments for which the pitchtrack was not reliable, though sometimes errors could be removed simply by minor adjustments of segment boundaries. The Emu interface to R was then used to query each database and extract a list of all target segments, along with their pitchtracks at 1 ms intervals. Then, values for the first and last 2% of each target interval were discarded to avoid F0 artifacts at segment edges. For the connected speech corpora, Emu was not used. Instead, F0 values at 10 ms intervals were output directly as text files and analyzed in Excel. The only checking and adjusting that was done, was done as part of examining and adjusting the automatically-generated Praat segmentations.

This VoiceSauce-based method thus was much more automated than the cepstral plus manual method, and less subject to pitchtracking failures, since STRAIGHT always returns an F0 value. With this method, no effort was made to find very low F0 values directly from waveforms. We will refer to this method as "semi-automated STRAIGHT".

Creaky voice is pervasive in our recordings of both languages. While many studies of F0, whether of F0 range or of intonation, ignore creaky intervals, it is important to understand the extent to which they affect the measures extracted from the recordings. Though not mentioned in previous cross-language F0 studies, creaky voice can cause particular problems for pitchtrackers. These difficulties could result in misleading estimates of speakers' minimum F0s, and thus affect estimates of F0 means and ranges. Therefore we include a comparison of manual vs. automated pitchtracking in creaky voice. In this way, the effect of relying on automated pitchtracking can be better understood.

2. F0 measures

Baken and Orlikoff (2000) review the basic and most common measures of F0, which include measures of the average F0 (usually the mean), of the F0 varaibility (usually the standard deviation, or "pitch sigma"), and of the overall F0 range (either the MaximumF0-MinimumF0, or some subset which removes outliers). To these measures, we add here measures of the most extreme F0 values produced by each speaker.

All the measures used in this study are summarized in Table II. Using one or both of the pitchtracking methods described in the previous section, the minimum (Min) and maximum (Max) F0 value was found for each sweep/vowel/passage (depending on the corpus). Where

there were multiple utterances from a speaker in a given corpus, these values were averaged, giving a mean minimum F0 and a mean maximum F0, which will be referred to as MeanMin and MeanMax. Also from multiple Min and Max values of a single speaker, F0 extremes were identified: the lowest minimum (XMin), and the highest maximum (XMax). Finally, for the STRAIGHT data only, the F0 values over time in each pitchtrack were used to calculate the mean and standard deviation for that track. Again, where there were multiple utterances, these are reported as averages (MeanMean and MeanSD).

Then in Excel, for data from both pitchtracking methods, four range measures were calculated from the minimum and maximum measures: (mean) range in Hz ((Mean)Max – (Mean)Min), the same in semi-tones, extreme range (XRange) in Hz (XMax-XMin) and the same in semi-tones.

TABLE II. F0 measures calculated.

Measure	Abbreviation	Unit	Definition
Minimum F0	Min	Hz	Lowest F0 value in a token
Maximum F0	Max	Hz	Highest F0 value in a token
Mean Minimum F0	MeanMin	Hz	Average Min across tokens
Extreme Minimum F0	XMin	Hz	Lowest Min across tokens
Mean Maximum F0	MeanMax	Hz	Average Max across tokens
Extreme Maximum F0	XMax	Hz	Highest Max across tokens
Mean F0 Range	MeanRange	Hz	MeanMax – MeanMin
Mean F0 Range in semi-tones	MeanRangeST	Semi-tone	39.863 * log(MeanMax/MeanMin)
Extreme F0 Range	XRange	Hz	XMax - XMin
Extreme F0 Range in semi-tones	XRangeST	Semi-tone	39.863 * log(XMax/XMin)
Mean F0	Mean	Hz	Average of F0 values in a token
Mean of Mean F0	MeanMean	Hz	Average Mean across tokens
Standard deviation of F0	SD	Hz	SD of F0 values in a token
Mean of Standard deviations of F0	MeanSD	Hz	Average SD across tokens

E. Statistical analysis

ANOVAs were performed on the F0 descriptive statistics described above, using either SPSS v.17, or the R interface to Emu. Specific tests will be described in the sections below.

III. ANALYSIS AND RESULTS

A. Prompted sweeps

We begin with the simplest corpus, the prompted sweeps, which providean established way of estimating a speaker's physiological F0 range. While Reich et al. (1990) reported that they eliminated F0 values for any portions perceived as vocal fry or falsetto, we included all F0

values as described below. For this corpus, the two pitchtracking methods described above were both applied, and compared. All the measures listed in Table II were made under the semi-automated STRAIGHT method, while Mean and SD measures were not made under the cepstral plus manual method. Measures of minimum, maximum, and mean F0, plus F0 range, were made separately for falling and rising sweeps. The extreme measures (XMin, etc.), in contrast, were necessarily calculated across falling and rising sweep data combined, in that XMin always comes from falling sweeps (since those sweeps reach the lowest pitches) while XMax always comes from rising sweeps (since those sweeps reach the highest pitches). F0 measures were compared in a series of 2-way and 3-way ANOVAs (using R or SPSS) in which between-subjects factors were Language (English, Mandarin) and Sex (Male, Female), while SweepType (Rising, Falling) was a within-subject (Repeated measures) factor in some analyses.

Not surprisingly, speaker Sex almost always has a significant effect on all F0 measures in Hz. Also not surprisingly, falling vs. rising sweeps generally have very different Min, Max and Range values. To avoid cluttering the statistical results, therefore, details of these significant effects will not be reported here. Only if they interact with the Language variable will they be discussed.

1. Cepstral plus manual measurements

Data from 44 speakers were included in these analyses (12 English women, 10 English men, 11 Mandarin women, 11 Mandarin men). The other two speakers were excluded because their voices screeched in a way that could not be pitchtracked. Speakers who used incorrect vowels or nasal consonants, or who used odd voice qualities, were nonetheless included in the dataset, as long as their pitches could be tracked for at least one token of each utterance type. Recall that with this method, all Min and Max values were checked by hand, including hand-measuring of Min values from waveforms, e.g. during creaky-voice intervals.

For this dataset, there were no main or interaction effects of Language on any measure. That is, Mandarin and English speakers were similar on all measures. The Sex and Sweeptype variables affected the measures as expected. Averaged across all speakers, the MeanMin of falling sweeps was 90 Hz, and the MeanMax of rising sweeps was 817 Hz. XMin was 79 Hz, XMax was 858 Hz, XRange was 779 Hz and 43 ST.

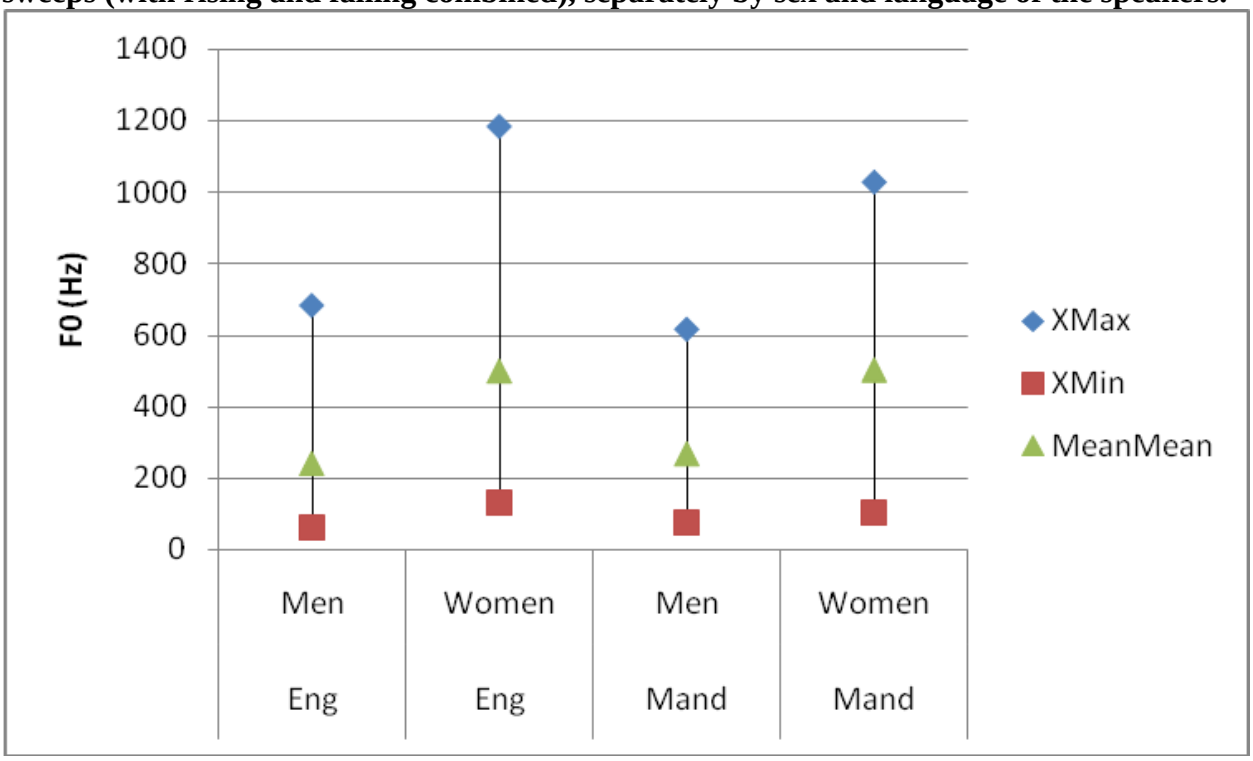
2. Semi-automated STRAIGHT

With this method, some manual checking was needed to eliminate gross errors of pitchtracking, as described above. Thus, creaky voice is sometimes included in analyzed segments (whenever the pitchtracker was successful), sometimes not (whenever the pitchtracker failed). As long as a speaker had at least one usable token of each utterance type, that speaker was included in the dataset. With these criteria, 41 of the 46 speakers were included. Two speakers (one English male, one English female) were excluded because of recording artifacts, and three speakers (all Mandarin females) were excluded because of untrackable voice qualities.

The measures and statistical analyses were otherwise the same as with the first method. Sex and Sweeptype affected the measures as expected and will not be further reported. The only (near-)significant effects of Language were seen in Sex x Language interactions for the two measures of Min F0. A significant interaction for MeanMin of falling sweeps ($F(1,37)=6.171$, $p=.018$) as well as a trend for XMin ($F(1,40)=3.938$, $p=.055$) both reflected relatively high Min

F0 values from the English women and relatively low values from the English men, such that the Sex difference was larger in English than in Mandarin. There were no significant Language differences within either Sex for MeanMin or XMin, though there were trends to significance for MeanMin. The three measures XMin, XMax, and MeanMean are shown by Language and Sex in Figure 1.

FIG. 1. Three F0 measurements, calculated by the STRAIGHT method, from the prompted sweeps (with rising and falling combined), separately by sex and language of the speakers.



For the measures compared by SweepType, there was a trend to a Language x Sweeptype interaction for MeanSD, but no significant main or interaction effects involving Language. Across all speakers, the average XRange was 774.2 Hz and 38.8 ST; the average MeanRange was 376 Hz and 19 ST, and the MeanMean was 375 Hz.

3. Comparison of results from the two methods

While the cepstral plus manual method produced measurements that did not show any effects of speakers' Language, the semi-automated STRAIGHT method found a significant difference in the size of the Sex effect in English vs. in Mandarin. Thus the two methods gave very similar results overall, with the semi-automated STRAIGHT method if anything more powerful.

B. Unprompted sweeps

Next we consider the unprompted sweeps, for which a special creaky condition was also elicited. The unprompted falling and rising sweeps were analyzed only with the semi-automated

STRAIGHT method, and these results are presented in this section. The unprompted creaky sweeps were also manually checked and will be presented in section C below. In this database, intervals with pitchtracking problems were excluded from most analysis (as before), but were included specifically in the measurement of XMin, the lowest observed F0 for a speaker. This was done because the goal was to include extreme values wherever possible, and based on visual inspection of tokens, such segments were likely to have spurious high F0 values (octave errors and the like), but did not seem to have spurious low F0 values. As long as a speaker had at least one usable token of each utterance type, that speaker was included in the dataset. With these criteria, measurements were obtained from 40 speakers (10 English women, 9 English men, 10 Mandarin women, 11 Mandarin men); 5 others had recording artifacts and one screeched his rising sweeps.

As above, the F0 measures for falling and rising sweeps were compared in a series of 2-way and 3-way ANOVAs (using SPSS) in which between-subjects factors were Language (English, Mandarin) and Sex (Male, Female), and SweepType (Rising, Falling) was a within-subject (repeated measures) factor. Again, results related to the Sex and Sweeptype factors apart from Language will not be reported.

There were no significant main or interaction effects with Language in any of the ANOVAs. There was a trend for Language to affect the MeanRangeST, seemingly due to English speakers having larger values in the rising sweeps, but this was not significant in a post-hoc test. The XRange measures, which here included Min F0s from creaked segments, showed no Language differences. Averaged across all speakers, XMin was 90.2 Hz, XMax was 706.3 Hz, and XRange was therefore 616.1 Hz and 36.6 ST; the MeanMin of falling sweeps was 117.9 Hz, the MeanMax of rising sweeps 677.7 Hz; the average MeanRange for falling and rising sweeps combined was 332.9 Hz and 19 ST; the MeanMean for falling and rising sweeps combined was 263.6 Hz and the MeanSD was 100 Hz.

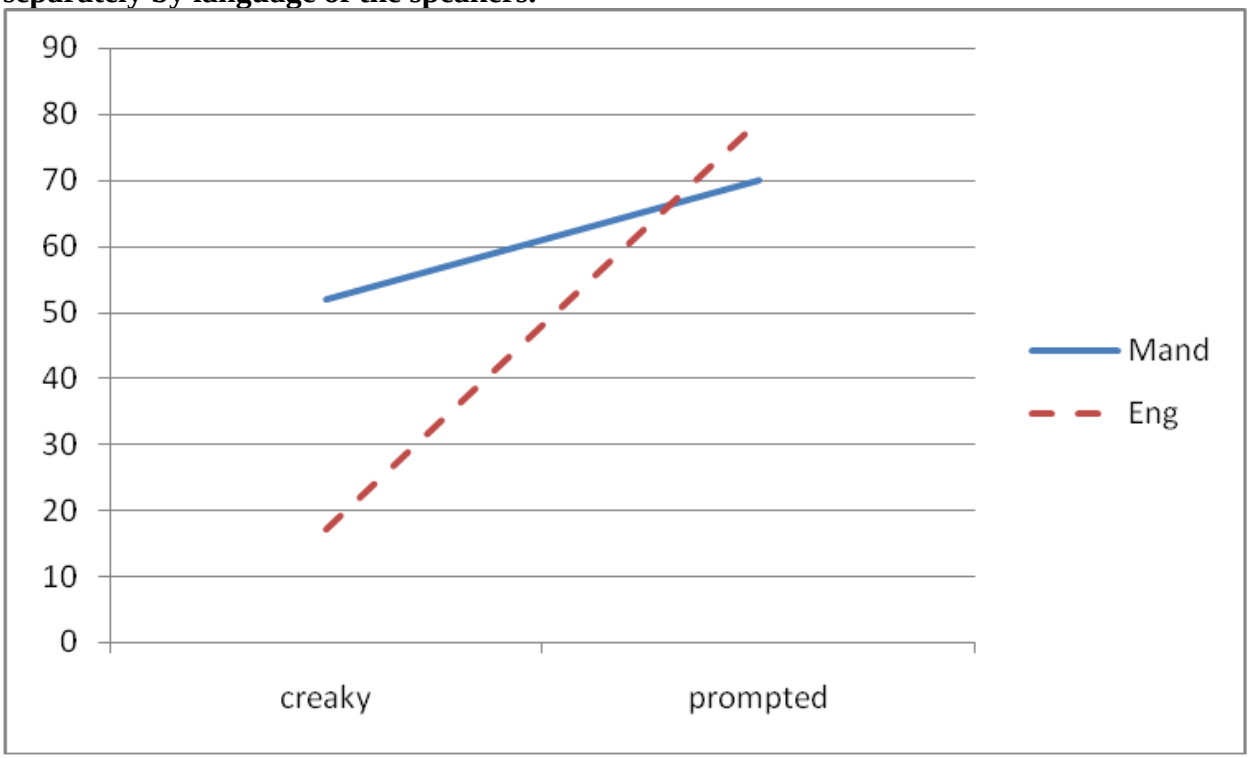
Another comparison that can be made concerns the starting pitches self-selected by the speakers in these unprompted sweeps. That is, do Mandarin and English speakers choose to begin a rise or a fall at similar pitches? Two-way ANOVAs (Language x Sex) were performed on the MeanMin for the rising sweeps and on the MeanMax for the falling sweeps. The pitches from the falling sweeps were the same across the two languages (263 Hz on average), but the pitches from the rising sweeps did differ between the languages ($F(1,36) = 5.0$, $p=.032$), with Mandarin speakers, both men and women, choosing higher starting pitches than the English speakers did (171 vs 145 Hz on average). And indeed, when just the rising sweeps are analyzed separately (rather than as part of 3-ways ANOVAs as above), the MeanMean does differ between the two languages, with Mandarin having the higher values (386 vs. 325 Hz); but the MeanRange measures do not differ.

C. Comparison of minimum F0 values

In the present study, for the prompted falling sweeps speakers were told nothing about how they might achieve the low pitches of the prompt. In contrast, as part of the unprompted sweeps corpus, speakers were asked to produce falling creaky sweeps, in which they allowed their voices to creak in order to reach the lowest possible pitch. Thus we can ask, in which of these two kinds of sweeps did the speakers of the present study produce lower pitches? We compare the XMin values from the creaky and prompted falling sweeps by our two analysis methods.

XMin values from the cepstral plus manual method (in which outputs were hand-checked against waveforms so that the Min for each utterance is accurate) were compared by a 3-way repeated-measures ANOVA, with Language and Sex as between-subjects variables and Sweeptype (unprompted creaky or prompted falling) as the within-subjects variable. The two Sweeptypes differed significantly ($F(1,38) = 46.68$, $p<.0001$), with the unprompted creaky having the lower XMin (34.8 vs 74.4 Hz). This main effect can be traced to two significant interaction effects, Sweeptype x Language ($F(1,38) = 14.563$, $p<.0001$) and Sweeptype x Language x Sex ($F(1,38) = 6.042$, $p=.019$), with everyone but Mandarin women having lower F0 in unprompted creaky than in prompted falling sweeps, and English speakers having the largest difference between the two sweeptypes. The Sweeptype x Language interaction is shown in Figure 2.

FIG. 2. Lowest F0 (XMin) measurement, calculated by the cepstral plus manual method, from the unprompted creaky vs. prompted falling sweeps (with the sexes combined), separately by language of the speakers.



The same analyses were repeated on the XMin values (for all segment labels) from the semi-automated STRAIGHT method. Here again the Sweeptypes were significantly different ($F(1,34)=85.473$, $p<.001$), but this time along with Sex ($F(1,34)=15.798$, $p<.001$). There was again no main effect of Language or Language x Sex interaction, but all the interactions with Sweeptype were significant, again including Sweeptype x Language x Sex ($F(1,34) =11.322$, $p=.002$) . Again, XMin from creaky sweeps was lower than from prompted falling (51.3 vs. 91.8 Hz), and now, as expected, males lower than females (56 vs. 87 Hz). The trend was for English men to have especially low creaky values and English women to have especially high prompted falling values.

The two pitchtracking methods give similar XMin results for the prompted falling sweeps, and under both methods XMin is lower in the unprompted creaky sweeps. Therefore we conclude that telling speakers to allow their voices to creak will, in general, result in lower measured Min F0 values. The lowest values are recovered only by hand-checking of individual tokens, and only with this method is there a significant interaction involving Language. Nonetheless, XMin values can be recovered reasonably well by the semi-automatic STRAIGHT method, and in the following analyses we will rely on this method.

D. Comparisons of results from sweeps

In sum, there are few ways in which speakers of the two languages differ in their physiological F0 ranges. Most notably, in the creaky sweeps when measured by hand, the Mandarin women's F0s do not go as low as other speakers', while in the prompted falling sweeps when measured by STRAIGHT, the English women's F0s do not go as low as other speakers'. In addition, the Mandarin speakers chose to begin their unprompted rising sweeps at higher F0s than the English speakers did. But there were no language differences in other Min, Max, Range, or Mean measures for either the prompted or the unprompted sweeps.

TABLE III: MeanMin, MeanMax, MeanRange across methods, and compared with 2 previous studies as cited by Baken & Orlikoff (2000)

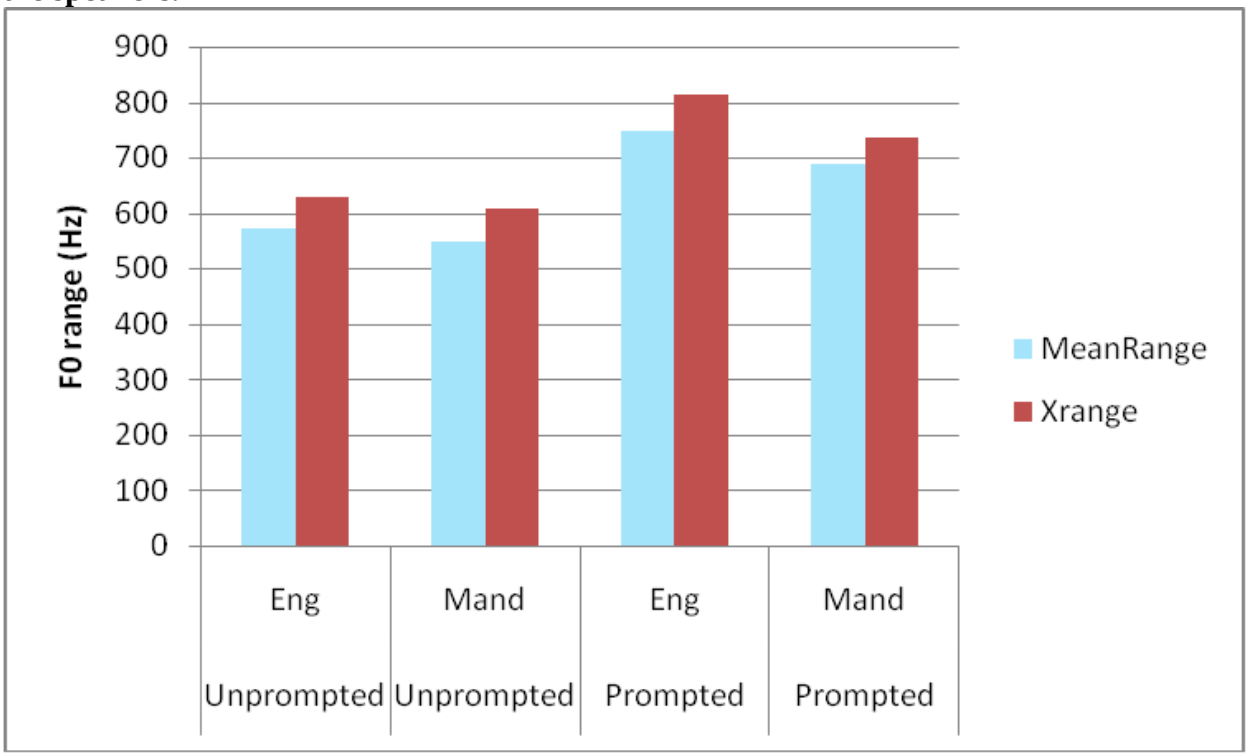
	Mean Min			Mean Max			Mean Range		
	men	women	both	men	women	both	men	women	both
Prompted cepstral	75	106	90	582	1052	817	507	946	727
Prompted STRAIGHT	77	136	106	595	1059	827	518	923	721
Unprompted STRAIGHT	23 (creaky) 85 (falling)	27 (creaky) 150 (falling)	25 (creaky) 118 (falling)	512	843	678	427	693	560
Hollien & Michel (1968)	24 (pulse) 94 (modal)	18 (pulse) 144 (modal)	21 119	634 (loft)	1131 (loft)	883	609	1113	861
Reich et al. (1990)	82	150	116	623	1125	874	541	975	758

Table III compares some of our measures across methods and studies. In our data, prompted sweeps gave larger ranges than unprompted (where the unprompted values come from rising and falling sweeps, except for MeanMin, where separate values are reported from English speakers' creaky sweeps) – see also Figure 3 below -- but smaller than results reported by Hollien and Michel (1968) (as shown in Table 6-20 in Baken and Orlikoff (2000)), or Reich et al. (1990) – even though the latter excluded perceived fry and falsetto speech from their

measurements. However, our values are more typical when compared to other previous studies summarized in Baken and Orlikoff's Table 6-16, where MeanRange for (non-elderly) men varies from 484 to 864 Hz, and for (non-elderly) women from 743 to 981 Hz. Our cepstral Min F0 measurements from the prompted sweeps include creaky intervals since these measurements were checked by hand as needed, and thus on average are lower than our STRAIGHT measurements. Our STRAIGHT Min measurements from the creaky sweeps were not checked by hand, and these are higher than Hollien and Michel's pulse values.

Previous studies did not include measures which were specifically about extreme F0 values. XMin, XMax, and XRange. XRange vs. MeanRange in our data are shown in Figure 3, which also graphically compares prompted vs. unprompted sweeps. It can be seen that not only are the ranges for prompted sweeps consistently larger than for unprompted, but XRange is, not surprisingly, consistently larger than MeanRange.

FIG. 3. Mean vs. extreme F0 ranges, calculated by the STRAIGHT method, separately by sweep type (unprompted vs. prompted, with rising and falling combined) and language of the speakers.



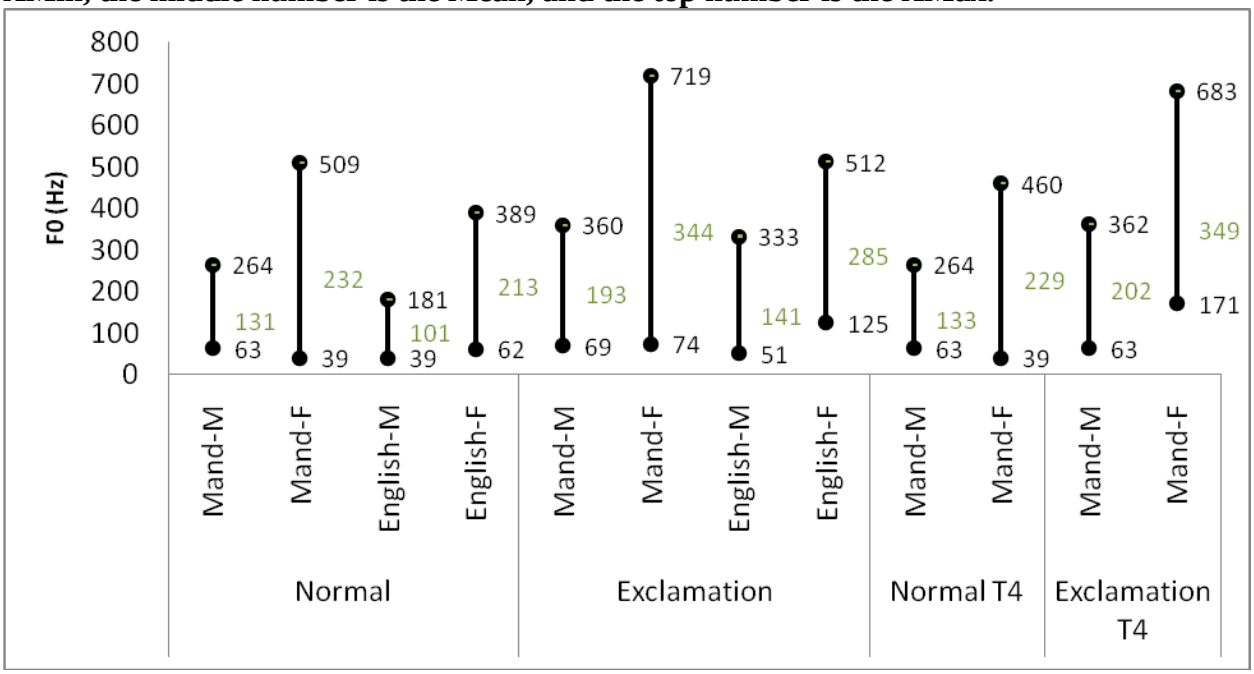
E. Isolated words

The sweeps corpora provide information about speakers' maximum physiological F0 range. The isolated words corpus, in contrast, provides information about speakers' use of F0 in their native languages. The part of this corpus analyzed here comprises productions of the English word *sure* and the 4 Mandarin "shi" words, all in two utterance types. For this corpus, measurements were made using both F0 analysis methods; however, as preliminary statistical tests indicated that the

two methods yielded similar results, the only analyses that will be reported here are those using the semi-automated STRAIGHT method, which permits more measures.

Utterances in this corpus were individually examined with their STRAIGHT pitchtracks, and the same corrective actions were taken as described above to prevent spurious/null F0 measurements. Two speakers (1 English woman, 1 Mandarin woman) were excluded for lack of usable tokens, giving data from 11 each of English men, English women, Mandarin men, and Mandarin women. Second, segment boundaries were slightly adjusted to put gross failures of tracking outside the target segments, especially to exclude F0 values of 0 (though also octave and other clear errors), but also to avoid spurious highs (especially likely at fricative edges) or lows. (Also note that removing the outside 2% of values on each edge also helped avoid spurious transitional F0 values.) However, as before, the pitch values themselves were never corrected or otherwise adjusted.

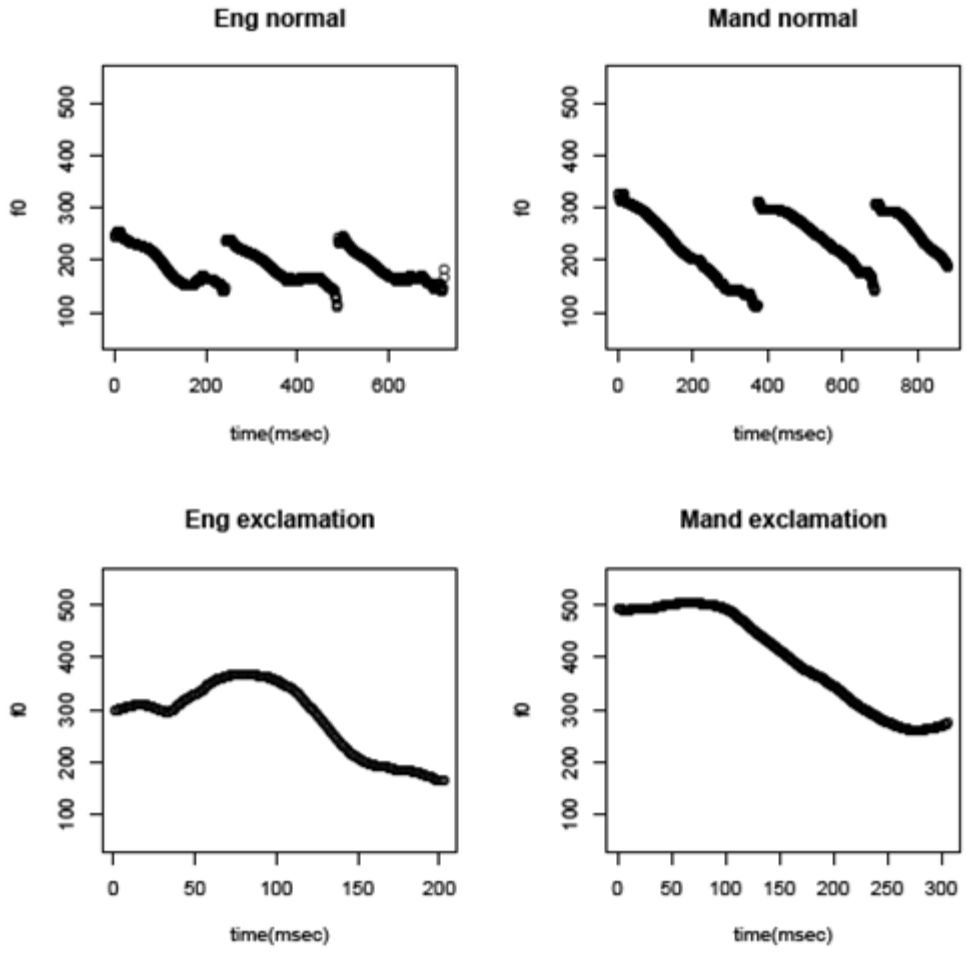
FIG. 4. Three F0 measurements, calculated by the STRAIGHT method, from the isolated words, separately by tone (all tones vs. only Tone 4), utterance type (normal pitch vs. exclamation), and language and sex of the speakers. On each bar, the bottom number is the XMin, the middle number is the Mean, and the top number is the XMax.



The most directly comparable utterances were selected for comparison: just the Mandarin Tone 4 utterances, which have a falling pitch contour that is most similar to the English utterances. Both normal pitch and exclamation utterances were compared; for the exclamation utterances, only the first token was selected, since that was the most comparable elicitation across the languages. Thus the measures here are within a single token for exclamations, but across repetitions for normal pitch utterances. Three-way ANOVAs used Language and Sex as between-subjects factors and UtteranceType as a within-subjects factor. In this analysis there were main effects of Language on four measures: Min (Mandarin 133 Hz vs. English 115 Hz, $F(1,40)=5.037$, $p=.03$), Max (Mandarin 321 Hz vs. English 258 Hz, $F(1,40)=16.162$, $p<.001$),

Range (in Hz) (Mandarin 189 Hz vs. English 143 Hz, F(1,40)=8.425, p=.006), and Mean (Mandarin 229 Hz vs. English 186 Hz, F(1,40)=18.383, p<.001). All of these differences were due to higher values in Mandarin than in English. That is, the Mandarin speakers did not go as low, but did go higher, with a greater range and mean. These results are shown in Figure 4, and sample utterances are given in Figure 5.

FIG. 5. Sample pitchtracks, calculated by the STRAIGHT method, of normal pitch vs. exclamation utterances from one English woman and one Mandarin woman (Tone 4 only).



In order to see the effect that lexical tones have on the language comparison, the same analyses were performed using the data from all four Mandarin tones (rather than just the falling Tone 4 as above). The results are very similar: Mandarin speakers had higher Max (332 vs. 258 Hz) and Mean (225 vs. 186 Hz) values, and greater ranges (221 vs. 143 Hz and 19 vs. 14 ST), and, in Exclamations only, higher Min values (140 vs. 124 Hz); however the English speakers had larger SDs (43 vs. 27 Hz). Some of these results are also shown in Figure 4.

That is, there is little difference in F0 measures between when Mandarin speakers produce all their tones, vs. just their falling Tone 4. The main exception is the standard deviation measure, which for Mandarin is much smaller in the all-tone data than in the Tone 4-only data, presumably due to the effect of the level tone, Tone 1. As a result, while the variability is the

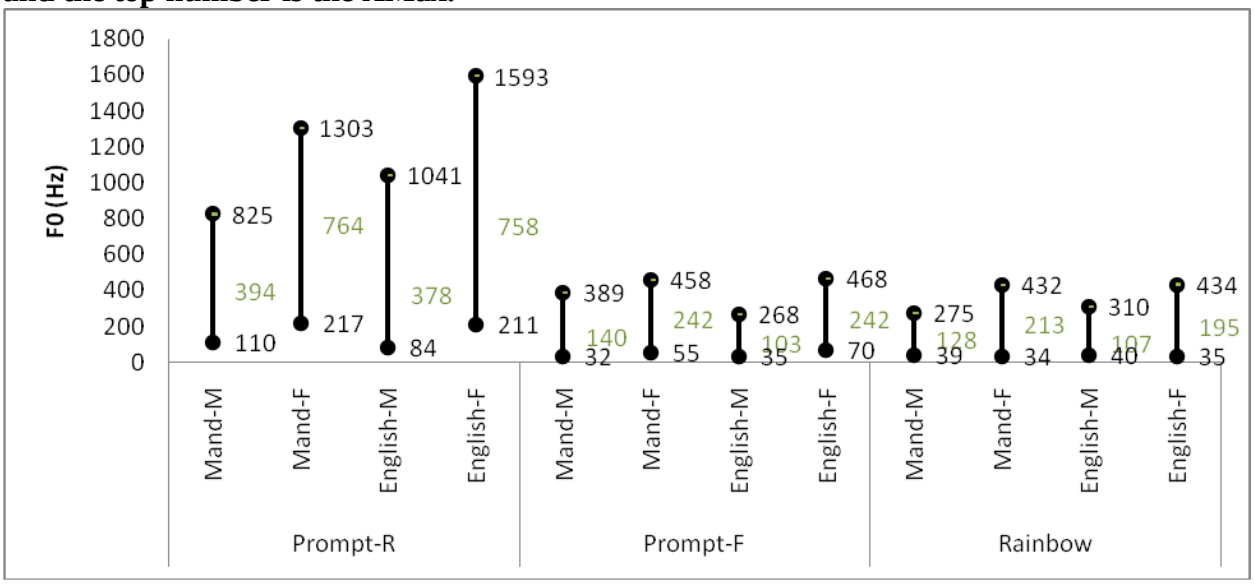
same for the two languages when only Tone 4 is included for Mandarin, overall variability is greater in English when compared to all the tones in Mandarin.

F. Rainbow passage

Like the isolated word corpus, the prose reading passage corpus provides information about speakers' use of F0 in their native languages, but in this case in connected reading. After exclusions for problems with recordings, passages from 18 Mandarin speakers (9 women, 9 men) and 18 English speakers (10 women, 8 men) were available for analysis. As before, Praat TextGrids were used to segment all the voiced portions of the recordings, and they were checked to prevent spurious or null F0 measurements, generally by removing the segment labels of problematic segments. That is, not all voiced segments were tracked in their entirety. The STRAIGHT F0 values of the pitchtracked segments were then output every 10 ms. (The typical recording yielded about 4800 such F0 values.) Descriptive statistics (Min, Max, Mean, SD, XRange, XRangeST) over each speaker's values were calculated in Excel and then analyzed as above by univariate ANOVAs, with Language and Sex as independent variables.

These analyses showed that, as expected, women had significantly higher Max, Mean, SD, and XRange, and also XRangeST; but the sexes did not differ on MinF0. There were no significant Sex x Language interactions. The two languages differed on only one measure, Mean ($F(1.32)=12.437$, $p=.001$), with Mandarin having the higher values (171 vs. 151 Hz). Thus, in this neutral reading passage, our Mandarin speakers used the same overall pitch ranges and variability as our English speakers, but they had higher average pitches within that range.

FIG. 6. Three F0 measurements, calculated by the STRAIGHT method, from the prompted sweeps (rising and falling) vs. the Rainbow passage, separately by language and sex of the speakers. On each bar, the bottom number is the XMin, the middle number is the Mean, and the top number is the XMax.



These results can be compared with those from the prompted sweeps, which tend to display speakers' widest physiological ranges. This comparison of average Min, Mean and Max F0 by language and sex is shown in Figure 6, with the rising and falling prompted sweeps given separately. The values for the falling prompted sweeps and the reading passage are strikingly similar. That is, in neutral reading, speakers on average approach their lowest pitches, and go about as high as the prompted starting pitches of the sweeps (specified by Reich et al. 1990).

G. Story dialog

Speakers read the story of Little Red Riding Hood twice, once neutrally and once using different voices for the dialog by the different characters in the story. From these second dialog readings, four intervals of speech were selected: (1) the conversation of Grandma with Wolf pretending to be Little Red Riding Hood; (2) an utterance by Wolf pretending to be Grandma (to Little Red Riding Hood); (3) the conversation of Little Red Riding Hood with Wolf pretending to be Grandma; (4) an utterance by Little Red Riding Hood (to the Woodsman). These intervals are indicated in the version of the story given in the Appendix and included both the character dialog and any narration that intervened. The pitchtracking and measuring method used for the Rainbow readings was applied to these intervals to yield the same set of F0 measures. As before, it cannot be guaranteed that the pitchtracks are error-free; especially in the stories, the very high pitches used by some speakers made the use of Max to screen for errors less reliable. After exclusions for problems with recordings, passages from 20 Mandarin speakers (10 women, 10 men) and 21 English speakers (11 women, 10 men) were available for analysis. As before, ANOVAs with Language and Sex as between-subjects variables were carried out on all the F0 measures.

These analyses showed that the women had significantly higher Min, Max, Mean, SD, and XRange (in Hz), but the sexes did not differ on XRangeST. There were no significant Language x Sex interactions. The two languages differed on most measures: Min, English 58 Hz vs. Mandarin 72 Hz ($F(1,37)=4.907$, $p=.033$), Max, English 627 Hz vs. Mandarin 490 Hz ($F(1,37)=12.512$, $p=.001$), SD, English 106 Hz vs. Mandarin 78 Hz ($F(1,37)=10.815$, $p=.002$), XRange, English 569 Hz vs. Mandarin 418 Hz ($F(1,37)=15.644$, $p<.001$), and XRangeST, English 41.3 Hz vs. Mandarin 33.5 Hz ($F(1,37)=17.834$, $p<.001$), but not for Mean, English 241 Hz vs. Mandarin 228 Hz ($F(1,37)=1.13$, $p=.295$). Many speakers, both English and Mandarin, used squeaky-high and/or growling low voices for one or more of the story characters. Overall, though, these statistical results indicate that English speakers of both sexes went both lower and higher than the Mandarin speakers, and thus had larger F0 ranges, in both Hz and ST, as well as larger F0 standard deviations. The Mean F0, however, did not differ between the two languages.

IV. DISCUSSION AND CONCLUSION

Here we summarize the results reported above, in the context of the study's research questions:

(1) Do English and Mandarin F0 profiles differ? In the pitch sweeps (prompted, unprompted), there were very few language differences. That is, the English and Mandarin speakers in this study appear to have the same physical capabilities with respect to rate of vocal fold vibration. In contrast, the speech samples showed several differences between the languages. Most often,

when there was a language difference, the Mandarin speakers used higher pitches and/or larger pitch ranges. Previous studies, mostly based on reading passages or other connected speech, have given a variety of results. Our reading passage result, that the average pitch is higher in Mandarin, agrees with Eady (1982) and Mang (2001). Like previous studies, we also found a greater pitch range for Mandarin, but only for the single word utterances. Note that in our data, standard deviation often does not pattern together with the range measures, indicating that the pattern of variability around the mean can be very different from the total extent of pitches produced, and thus one measure cannot be substituted for the other.

Crucially, the different speech samples showed different patterns of results. That is, whether the two languages will appear to have similar or different F0 profiles very much depends on the speech corpus or task. The inconsistency of results from previous studies may be partly due to differences in materials chosen.

(2) How does the type of voice sample elicited affect the F0 characterization? As just noted, physiological vs. speech comparisons showed very different results, and even within the speech samples, quite opposing differences were found across corpora. For single-word utterances (whether neutral or exclamatory, all Mandarin tones or Tone 4 only), the Mandarin speakers had higher values on most measures. However, for the Rainbow passage reading, which was neutral in style, there was no difference in the F0 ranges or standard deviations, but the Mandarin Mean was higher. Thus the two languages appear most similar in the prose passage reading. Finally, for the story voices in Little Red Riding Hood, quite the opposite result obtained: the Means were the same between the languages, but for the other measures the English values were more extreme.

Comparing the lively story voices to the exclamatory single words (with all four Mandarin tones), virtually all the measures are more extreme in both languages for the story voices, except that the Mean is the same in Mandarin. There seems to be a tendency for Mandarin speakers to maintain a consistent mean speaking F0.

Comparison of the F0 characteristics of the Rainbow passage readings vs. the prompted falling sweeps showed that within each language they are similar. That is, these two kinds of pitch sample seem to give essentially the same information about voices—how low a speaker can go, and how high is comfortable for non-expressive speech—across the languages.

(3) Can any differences be related to the fact that Mandarin is a lexical tone language? In the single-word utterances, English falling intonation contours are directly compared with the four Mandarin lexical tones. It is this corpus that shows the most difference between the two languages. Even when only the Mandarin high-falling Tone 4 is compared with the falling intonation contour in English—when the overall pitch contours are most similar between the languages—the differences are robust. Thus the language differences cannot be due to the fact that Mandarin has multiple lexical tones, or that it has a high level tone. Instead, the differences must be due to the way Mandarin speakers pronounce the high-falling tone, reaching a higher pitch peak and giving an overall larger pitch range and higher average pitch.

For the prose passage readings, the only difference between the languages is in the Mean F0, which was higher in Mandarin. The Mandarin speakers, though they covered the same F0 range as the English speakers, spoke about 20 Hz higher on average. Could the occurrence of high level Tone 1s in the text be responsible for this difference? This seems unlikely, since in this text there are relatively few Tone 1s, and further, the distribution is similar in the Little Red

Riding Hood selection, where the Mean F0 did not differ between the languages. Similarly, possible segmental influences from pitch-raising voiceless consonants, seem very brief in extent and unlikely to differ greatly between these language samples. However, it is possible that some combination of such small effects – the pitch peaks from the natural distribution of tones 1 and 4, plus segmental influences – could account for a 20 Hz difference in the means in the passages, but controlled studies would be needed to establish this.

Our results provide some support for the hypothesis that a tone language like Mandarin can have an overall higher average pitch, as this is what was found with both the single words and the prose passage. They also provide some, albeit weaker, support for the hypothesis that a tone language can have an overall larger pitch range, as this is what was found with the single words, though not with the prose passage. The Tone 4 data from the single word comparisons indicates that this tone alone can produce the observed language differences.

(4) How does the method of F0 estimation affect the F0 characterization? Differences in results with the cepstral plus manual vs. semi-automated STRAIGHT methods were small, and we concluded that they did not justify the much greater effort required for the cepstral-plus-manual method, which especially for the connected prose passages would have been extremely time-consuming, and which could not provide as many kinds of measurements. However, it is clear that the lowest creaky F0 values will be better detected manually.

In conclusion, we found differences in the F0 profiles of Mandarin vs. English speakers, but these differences depended on the particular speech samples being compared. Most notably, the physiological F0 ranges of the speakers, as determined from tone sweeps, did not differ between the two languages, indicating that the speakers' voices are comparable. Their use of voice pitch in single-word utterances was, however, quite different, with the Mandarin speakers having higher maximums and means, and larger ranges, even when only the Mandarin high falling tone was compared with English. In contrast, in a prose passage, the two languages were more similar, differing only in the mean F0, Mandarin again being higher. Finally, however, in a lively reading of story character voices, the English speakers had the higher maximums and means and the larger ranges.

The study thus contributes to the growing literature showing that languages can differ in their F0 profile, but highlights the fact that the choice of speech materials to compare can be critical.

ACKNOWLEDGMENTS

This research was supported by NSF grant BCS-0720304 to the first author. A very preliminary version was presented at the Spring 2009 meeting of the Acoustical Society of America in Portland. We thank Yen-Liang Shue for all his help with VoiceSauce, and undergraduate research assistants Ting Fan, Spencer Lin, Larina Luu, and Caitlin Smith for help with recording and analysis.

REFERENCES

Baken, R., and Orlikoff, R. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular Publishing, San Diego), pp. 145-223.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**, 341-345.

Chen, G.T. (**1972**). "A comparative study of pitch range of native speakers of Midwestern English and Mandarin Chinese: An acoustic study," doctoral dissertation, U. Wisconsin-Madison, Madison.

Chen, S. (**2005**). "The effects of tones on speaking fundamental frequency and intensity ranges in Mandarin and Min dialects," J. Acoust. Soc. Am. **117**, 3225-3230

Crystal, D. (**1969**). *Prosodic Systems and Intonation in English* (Cambridge U. P., London), pp. 126-194.

Deutsch, D., Jinghong, L., Sheng, J., and Henthorn, T. (**2009**). "The pitch levels of female speech in two Chinese villages," J. Acoust. Soc. Am. **125**, 208-213.

Dolson, M. (**1994**). "The pitch of speech as a function of linguistic community," Music Percept. **11**, 321-331.

Eady, S. J. (**1982**). "Differences in the F0 patterns of speech*:* Tone language versus stress language," Lang Speech **25**, 29-42.

Fairbanks, G. (**1940**). *Voice and Articulation Drill Book* (Harper, New York), p. 168.

Fairbanks, G. (**1960**). *Voice and Articulation Drill Book*, 2nd ed. (Harper, New York), pp. 124-139.

Graddol, D., and Swann, J. (**1989**). *Gender Voices* (Blackwell Publishing, Oxford), pp. 12-40.

Graddol, D., and Swann. J. (**1983**). "Speaking fundamental frequency: Some physical and social correlates," Lang Speech **26**, 351-366.

Hanley, T.D., and Snidecor, J.C. (**1967**). "Some acoustic similarities among languages," Phonetica **17**, 141-148

Hanley, T.D., Snidecor, J.C., and Ringel, R. (**1966**). "Some acoustic differences among languages," Phonetica **14**, 97-107.

Harrington, J. (**2010**). *Phonetic Analysis of Speech Corpora* (John Wiley and Sons, West Sussex), Chap. 2.

Henton, C. (**1989**). "Fact and fiction in the description of female and male pitch," Language and Communication **9**, 299–311.

Hollien, H., and Michel, J.F. (**1968**). "Vocal fry as a phonational register," J. Speech Hear. Res. **11**, 600-604.

Honorof, D.N., and Whalen, D.H. (**2005**). "Perception of pitch location within a speaker's F0 range," J. Acoust. Soc. Am. **117**, 2193–2200.

Kawahara, H., de Cheveign, A., and Patterson, R. D. (**1998**). "An instantaneous-frequency-based pitch extraction method for high quality speech transformation: revised TEMPO in the STRAIGHT-suite," Proc. ICSLP'98, Sydney, Australia, December 1998.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge), pp. 68-105

Liu, H.-M. (**2002**). "The acoustic-phonetic characteristics of infant-directed speech in Mandarin Chinese and their relation to infant speech perception in the first year of life," doctoral dissertation, U. Washington.

Loveday. (**1981**). "Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae," Lang Speech **24**, 71-89.

Majewski, W., Hollien, H., and Zalewski, J. (**1972**). "Speaking fundamental frequency of Polish adult males," Phonetica **25**, 119-125.

Mang, E. (**2001**). "A cross-language comparison of preschool children's vocal fundamental frequency in speech and song production," Research Studies in Music Education **16**, 4-14.

Mennen, I., Schaeffler, F., and Docherty, G. (**2007**). "Pitching it differently: A comparison of the pitch ranges of German and English speakers," 16th Int. Congress of Phonetic Sciences, Saarland University, Saarbrücken.

Ohara, Y. (**1992**). "Gender-dependent pitch levels: A comparative study in Japanese and English," *Locating power: Proceedings of the second Berkeley women and language conference*, edited by K. Hall, M. Bucholtz, and B. Moonwoman (Berkeley: Berkeley Women and Language Group, Berkeley), pp. 469-477.

Reich, A.R., Frederickson, R.R., Mason, J.A., and Schlauch, R.S. (**1990**). "Methodological variables affecting phonational frequency range in adults," J. Speech Hear. Disord. **55**, 124-131.

Shue,Y.-L., Keating, P., and Vicenik, C. (**2009**). "VOICESAUCE: A program for voice analysis," J. Acoust. Soc. Am. **126**, 2221 (A)

Todaka, Y. (**1993**). "A cross-language study of voice quality: bilingual Japanese and American speakers," doctoral dissertation, UCLA, 145-147.

Torgerson, R.C. (**2005**). "A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of three native speech styles," master's thesis, Brigham Young U., 73-82

Yamazawa, H., and Hollien, H. (**1992**). "Speaking fundamental frequency patterns of Japanese women," Phonetica **49**, 128-140.

Xue, A., Hagstrom, F. and Hao, G. (**2002**). "Speaking F0 characteristics of bilingual Chinese-English speakers: A functional system approach," Asian Pacific J. Speech, Language and Hearing **7**, 55-62.

APPENDIX

Portions of the text of the story of Little Red Riding Hood used in this study – beginning and end not reproduced here. Bold-faced text was analyzed in the present study.

While little Red Riding-Hood was playing in the wood, the great wolf galloped on as fast as he could to the old lady's house. Now, grandmother was very feeble, and it happened that she was in bed that day. When the wolf reached the cottage door he tapped.

"Who is there?" asked the old lady.

"Little Red Riding-Hood, grandmother," said the wolf, trying to speak like the child. "Come in, my dear," said the old lady, who was a little deaf. "Pull the string and the latch will come up."

The wolf did as she told him, and went in, and you may think how frightened poor grandmother was when she saw him instead of Little Red Riding-Hood.

Now, the wolf, who was quite hungry after his run, soon ate up the poor old lady. Indeed, she was not enough for his breakfast, and so he thought he would like to eat sweet Little Red Riding-Hood also. Therefore, he dressed himself in grandmother's night-cap and got into bed, and waited for the child to knock at the door.

By-and-by, Little Red Riding-Hood reached her grandmother's house, and tapped at the door.

"Come in," said the wolf, in a squeaking voice. "Pull the string, and the latch will come up."

Little Red Riding-Hood thought her grandmother must have a cold, as she spoke so hoarsely; but she went in at once, and there lay her grandmother, as she thought, in bed.

"If you please, grandmother, mother has sent me with some blackberry wine."

But when Little Red Riding-Hood saw the wolf she felt frightened. She had nearly forgotten her grandmother, but she did not think she had been so ugly.

"Grandmother," she said, "what a great nose you have."

"All the better to smell with, my dear," said the wolf.

"And, grandmother, what large ears you have."

"All the better to hear with, my dear."

"Ah! grandmother, and what large eyes you have."

"All the better to see with, my dear," said the wolf, showing his teeth, for he longed to eat the child up.

"Oh, grandmother, and what great teeth you have!" said Little Red Riding-Hood.

"All the better to eat you up with," growled the wolf, and, jumping out of bed, he rushed at Little Red Riding-Hood, and would have eaten her up, but just at that minute the door flew open, and a great dog tore him down. The wolf and the dog were still fighting when Hugh, the woodman, came in and killed the wicked wolf with his axe.

Little Red Riding-Hood threw her arms round the woodman's neck, and thanked him again and again.

"Oh, you good, kind Hugh, she said, how did you know the wolf was here, in time to save me?"