ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Hearing Research 224 (2007) 75–83

Hearing Research

www.elsevier.com/locate/heares

Research paper

The perception of FM sweeps by Chinese and English listeners

Huan Luo [a,b,*], Anthony Boemio [d], Michael Gordon [b], David Poeppel [a,b,c]

[a] Neuroscience and Cognitive Science Program, University of Maryland College Park, 1401 Marie Mount Hall, College Park, MD 20742, USA
[b] Department of Biology, University of Maryland College Park, USA
[c] Department of Linguistics, University of Maryland College Park, USA
[d] Laboratory of Brain and Cognition, National Institute of Mental Health, USA

Received 3 August 2006; received in revised form 9 November 2006; accepted 17 November 2006

Abstract

Frequency-modulated (FM) signals are an integral acoustic component of ecologically natural sounds and are analyzed effectively in the auditory systems of humans and animals. Linearly frequency-modulated tone sweeps were used here to evaluate two questions. First, how rapid a sweep can listeners accurately perceive? Second, is there an effect of native language insofar as the language (phonology) is differentially associated with processing of FM signals? Speakers of English and Mandarin Chinese were tested to evaluate whether being a speaker of a tone language altered the perceptual identification of non-speech tone sweeps. In two psychophysical studies, we demonstrate that Chinese subjects perform better than English subjects in FM direction identification, but not in an FM discrimination task, in which English and Chinese speakers show similar detection thresholds of approximately 20 ms duration. We suggest that the better FM direction identification in Chinese subjects is related to their experience with FM direction analysis in the tone-language environment, even though supra-segmental tonal variation occurs over a longer time scale. Furthermore, the observed common discrimination temporal threshold across two language groups supports the conjecture that processing auditory signals at durations of ∼20 ms constitutes a fundamental auditory perceptual threshold.

Keywords: Tone language; Temporal threshold; FM identification; FM discrimination; Signal detection; Bias; FM direction selectivity

1. Introduction

Frequency modulation (FM) is an important physical aspect of communication sounds in both human and other species (Doupe and Kuhl, 1999). That the auditory system, in general, is well equipped to analyze FM signals has been shown in a range of neurophysiological studies in animals (Whitfield and Evans, 1965; Suga, 1968; Heil et al., 1992; Mendelson et al., 1993; Eggermont, 1994, 2001; Liang et al., 2002) as well as in human psychophysical (Schouten, 1985, 1986; Dooley and Moore, 1988; Schouten and Pols, 1989; Edwards and Viemeister, 1994a,b, 1997; Madden and Fire, 1997; Moore, 1997; Moore and Sek, 1992; Gordon and Poeppel, 2002) and brain imaging studies (Makela et al., 1987; Dimitrijevic et al., 2001; John et al., 2001, 2002; Picton et al., 1987, 2003; Boemio et al., 2005; Luo et al., 2006).

Tone sweeps, or FM glides, a special case of frequency-modulated signals, constitute a fundamental acoustic component of many complex sounds, particularly speech sounds. This is particularly true on the shorter, 'local' time scale ranging from 20 to 80 ms (e.g., formant transitions characteristic of consonants, Liberman et al., 1956) and on the longer 'global' scale from 200 to 300 ms (e.g., prosody and melody, phenomena at the syllabic rate, cf. Poeppel, 2003). Importantly, supra-segmental FM forms the basis for the lexical distinctions in tone languages, such

---

Abbreviations: FM, frequency modulated; SDT, signal detection theory; d', *d* prime

[*] Corresponding author. Address: Neuroscience and Cognitive Science Program, University of Maryland College Park, 1401 Marie Mount Hall, College Park, MD 20742, USA. Tel.: +1 301 405 2587; fax: +1 301 314 5052.

E-mail address: luohuan@gmail.com (H. Luo).

0378-5955/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2006.11.007

as Mandarin Chinese or Thai, both of which use highly constrained and parameterized pitch contour variations to make lexical distinctions (Howie, 1976; Stagray et al., 1992). For example, the segmental sequence, [ma], has four distinct meanings in Mandarin Chinese when spoken with different supra-segmental (or 'global') pitch contours. In contrast, for the English listener there is just a single lexical-semantic interpretation, regardless of the possible acoustic variation. Therefore, psychophysical investigation of the sensitivity of the human auditory system to FM sweeps can help elucidate the processing of speech in the human brain. That is not to say, of course, that the relation between processing FM signals and speech is straightforward. Speech sounds have a highly complex spectro-temporal structure, and the extent to which the proficiency in processing FM signals predicts speech discrimination is not yet fully understood. Nevertheless, a basic understanding of the thresholds associated with the perceptual analysis of FM glides sheds light on how the auditory system deals with stimuli that incorporate such transitions.

One issue reflecting emerging consensus is that there exist relatively clear rate/duration thresholds for human listeners below which the robust identification of FM direction becomes problematic. For example, both Schouten (1985) and Gordon and Poeppel (2002) observed that UP FM glides require at least 20 ms duration to identify direction. On the other hand, there is more disagreement concerning FM directional sensitivity. Data from psychophysics, animal physiology and human imaging data have found all the possible patterns: UP better than DOWN (Schouten, 1985; Madden et al., 1997; Gordon and Poeppel, 2002; Maiste and Picton, 1989; Rupp et al., 2002; Dau et al., 2000), DOWN better than UP (Dooley and Moore, 1988; Heil et al., 1992; Eggermont, 2001), and no directional sensitivity difference (Moore and Sek, 1992).

The general question evaluated in the present study is whether native language environment affects performance on identification and discrimination of non-speech signals containing basic (speech-related) FM features. We are testing the hypothesis that language experience conditions psychophysical performance on non-speech auditory tasks. Stated in the context of a more neurobiologically centered discussion, we are evaluating whether language-related experience-dependent plasticity extends to other acoustic domains. In fact, there is evidence, for example from PET studies (Gandour et al., 1998, 2000, 2002), indicating that subjects engage different neural circuitry when they perform low-level acoustic tasks if those sounds are relevant to their language. Specifically, speakers of tone languages must attend to the FM cues in speech signals to permit successful lexical access, leading to the hypothesis that the analysis of such acoustic information is more developed in the speaker of a tone language. Indeed, speakers of tone languages might show lower thresholds for such signal detection and better performance in the FM identification task, if extensive processing expertise at one time-scale (slow pitch contour) can transfer to perceptual analysis at another time scale (fast FM). On the other hand, there should exists a relatively similar 'basic' temporal threshold across different language groups, deriving from biophysical properties of auditory neuronal ensembles (in cortex), from which people can process and recognize a variety of sounds in their environment, including speech, music, and other natural sounds. These two seemingly paradoxical hypotheses can be examined by changing experimental procedures and design, even employing the same class of stimuli on the same subjects. Here we employed two experimental designs: FM *identification* and FM *discrimination*.

In Experiment 1, the FM direction identification experiment, Mandarin Chinese speakers were tested with two of the three stimulus sets used in Gordon and Poeppel (2002) on English speakers, with the frequency span comparable to the first and third formant ranges in speech sounds (i.e., 600–900 Hz and 2–3 kHz). We held constant the frequency ranges and parameterized FM rate by varying FM duration. The purpose of this experiment was to find out whether tone-language speakers perform better and manifest lower thresholds in this FM identification experiment. In addition, we asked the question whether they show similar direction sensitivity, as demonstrated in English subjects in previous experiments (Gordon and Poeppel, 2002). In Experiment 2, the FM discrimination experiment, both English and Chinese groups were tested. The goal here was to investigate whether the two groups manifest similar performance in a basic FM discrimination task. Note that to accomplish the discrimination task, subjects only needed to detect differences between FM sounds of opposite directions rather than explicitly identifying directions, as required in Experiment 1, and thus can depend on other acoustic properties or more holistic cues. By combining the results from the two experiments, we provide further evidence for a basic human temporal processing threshold at approximately 20 ms, and also identify a possible influence deriving from native language experience.

2. Materials and methods

2.1. Experiment 1 (FM identification)

2.1.1. Participants

Twelve Mandarin Chinese speakers [3F, age 20–35 years], with normal hearing, recruited from the University of Maryland student and staff population, participated in the experiment after providing informed consent. Subjects had no history of hearing or neurological problems.

2.1.2. Stimuli

The stimuli were linearly frequency-modulated tone sweeps, generated with Matlab (Mathworks, Natick, MA) with 16 bits of resolution and sampled at 44.1 kHz.

Both upward (UP) and downward (DOWN) FM sweeps at 10 different rates of frequency change (FM rate) were presented in each of two frequency ranges: 600–900 Hz and 2–3 kHz. The bandwidth of all FM stimuli was kept constant at half an octave in order to approximate the bandwidth of formant transitions. For each frequency range tested, the stimulus set comprised 20 different FM sweeps (2 directions × 10 FM rates). FM rate was parameterized by varying the duration of the signals. The following FM durations [with corresponding FM rates] were used: 5 ms [100 oct/s], 10 ms [50 oct/s], 20 ms [25 oct/s], 30 ms [16.7 oct/s], 40 ms [12.5 oct/s], 50 ms [10 oct/s], 80 ms [6.2 oct/s], 160 ms [3.1 oct/s], 320 ms [1.6 oct/s], and 640 ms [0.8 oct/s]. The relative intensity of the stimuli was adjusted to compensate for the duration-intensity trade-off, i.e., to make all stimuli of roughly equal loudness. All FM stimuli had a linear rise and fall time of 2 ms to minimize spectral splatter.

2.1.3. Procedure

Each of the two frequency ranges was tested in separate experimental blocks. In a single-interval two-alternative-forced-choice (2AFC) task, subjects had to identify the direction of FM (UP versus DOWN) by pressing one of two labeled keys. After half of the trials, we reversed keys to eliminate response key bias. Before each experiment, subjects were given a brief practice session to familiarize them with the stimuli. The experiment consisted of 400 trials (20 repetitions per stimulus condition) presented in pseudo-random order using PsyScope (CMU, Pittsburgh, USA) on a Macintosh system. The inter-trial interval was varied between 750 ms and 1750 ms. Subjects were given a short rest after the first 200 trials. All stimuli were presented binaurally at a comfortable loudness level.

2.1.4. Data analysis

Percent correct performance was calculated across 20 trials for each stimulus condition and for UP and DOWN FM stimuli separately, resulting in two groups of data, corresponding to UP FM identification (UP score) and DOWN FM identification (DOWN score) performance. Because the human perceptual system can be viewed as a binary signal detection system (UP or DOWN) during this 2AFC FM identification experiment, we further quantified the performance using Signal Detection Theory (SDT) by calculating *d*-prime, representing detector sensitivity, and $\beta_{\text{normalized}}$, representing subject bias (Rosenblith and Stevens, 1953; Dorfman and Alf, 1968; Swets, 1982). Specifically, the Hit and False Alarm (FA) rates and corresponding *d*-prime (d') and $\beta_{\text{normalized}}$ were defined as follows:

$$
\begin{aligned}
&Hit = H(UP\ pressing/UP\ stimulus) = UP\ score\\
&FA = 1 - H(Down\ pressing/DOWN\ stimulus)\\
&\quad = 1 - DOWN\ score\\
&d' = Zscore(Hit) - Zscore(FA)\\
&\beta = -0.5*(Zscore(Hit) + Zscore(FA))
\end{aligned}
$$

$$
\begin{aligned}
\beta_{normalized} = \beta/d' &= 0 \quad no\ bias\\
&< 0 \quad UP\ bias\\
&> 0 \quad DOWN\ bias
\end{aligned}
$$

With the introduction of $\beta_{\text{normalized}}$, we can detect possible bias strategies that subjects may employ during circumstances in which they cannot easily identify stimulus direction. For example, if a subject were to repeatedly press the 'UP' button during the identification task, his/her performance would be (artificially) high while $\beta_{\text{normalized}}$ would be negative indicating a "press-UP" bias strategy.

2.2. Experiment 2 (FM discrimination)

2.2.1. Participants

Six English subjects and six Chinese subjects with normal hearing participated in this experiment after providing informed consent. Subjects had no history of hearing or neurological problems.

2.2.2. Stimuli

Experiment 2 used all 10 stimuli from Experiment 1 except 640 ms. Because the pattern of results was similar for both frequency ranges used in Experiment 1, only the 2–3 kHz frequency range was examined in Experiment 2.

2.2.3. Procedure

A three-interval, two-alternative-forced-choice (3IFC) task, with three intervals per trial, each containing one sound was utilized to examine the FM discrimination performance. The inter-interval time was set at 500 ms. The first interval contained a 'standard' FM sound with direction either UP or DOWN. In the following two intervals, one contained exactly the same sound as the 'standard' and the other contained an FM sound of same duration but opposite direction. Subjects had to choose which of the two intervals following the standard contained the different sound. There were 20 trials for each FM duration. All stimuli were presented binaurally at a comfortable loudness level. The inter-trial interval was varied between 750 ms and 1750 ms (pseudo-randomly). Subjects were asked to make the decision as accurately as possible.

2.2.4. Data analysis

Percent correct scores were calculated over all 20 trials for each FM duration regardless of the exact stimulus configuration in each trial. This produced one data set for each subject, expressing FM discrimination ability as a function of rate (manipulated here as a change in FM duration from 320 ms to 5 ms, with a constant frequency span of half an octave).

2.3. Frequency control experiment

An additional experimental block was created in which we pseudo-randomly interleaved UP and DOWN trials from three frequency ranges (600–900 Hz, 1–1.5 kHz, 2–3 kHz) to eliminate the possible frequency cues by randomizing FM starting (and ending) frequencies. In previous experiments, the two frequency ranges (600–900 Hz, 2–3 kHz) were tested in separate blocks and therefore the two types of FM sweep stimuli (UP and DOWN) in each block were also different in their starting frequency in addition to the FM sweep direction. Specifically, UP FM stimuli had low starting frequency, and DOWN FM stimuli had high starting frequency, and starting frequency could be possibly be used as an acoustic cue in the identification task. Two of the same subjects that participated in Experiments 1 and 2 (one English, one Chinese) were run on this control task.

3. Results

3.1. Experiment 1 (2AFC)

Fig. 1 shows the Chinese participants' percent correct performance for the 600–900 Hz range (a) and the 2–3 kHz range (b). These data are compared directly to the data from Gordon and Poeppel (2002) in Fig. 1c and d, which show the comparable ratings for English speakers tested with the same stimuli and the same experimental procedure. The data for the Chinese subjects show that performance was at ceiling for both UP and DOWN FM directions when duration was above 20 ms (nearly 98% correct). When duration was very short (highest FM rate at signal durations of 5 ms and 10 ms), there was an asymmetry between UP and DOWN FM identification performance. The performance for DOWN FM identification remained high even at 5 ms signal duration, but the performance for UP FM identification dropped quickly and below chance level when FM duration was below 20 ms. An analysis of variance testing *direction* and *duration/FM rate* on the 2–3 kHz frequency range showed a main effect of *duration/FM Rate* [$F(9,99) = 30.30$, $p < 0.001$], a main effect of direction [$F(1,11) = 14.14$, $p < 0.001$], and a significant direction-by-duration interaction [$F(9,99) = 8.15$, $p < 0.001$]. Further post hoc analysis of duration effects showed that performance for 5 ms and 10 ms signals was

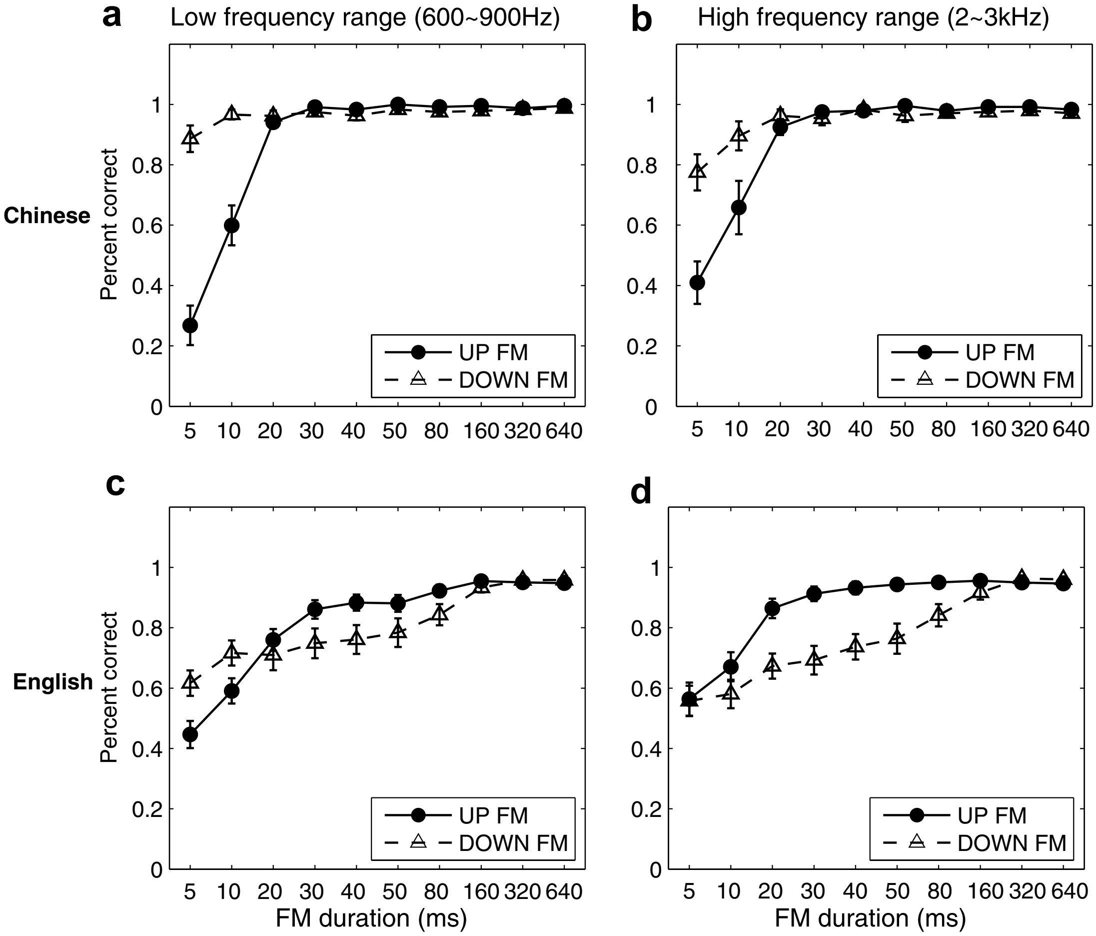

Fig. 1. FM direction identification results for Mandarin Chinese (a, b) and English listeners (c, d) on the single-interval two-alternative forced-choice (2AFC) paradigm. On each trial, participants heard a single linear FM sweep of fixed bandwidth (600–900 Hz (a, c); 2–3 kHz (b, d)) and duration as indicated along the *x*-axis (5–640 ms, corresponding to 100–0.8 oct/s). Subjects indicated by button-press whether the FM sweep direction was UP or DOWN, and the UP (solid line) and DOWN (dotted line) FM identification percent correct scores are shown in the same figure. Error bars are SEM.

significantly different from the longer stimuli; there were not significant differences between the other durations. (Because in our experiment FM rate maps to duration one-to-one, we will use duration to represent duration/FM rate in the following parts.) The reaction time for UP and DOWN did not show any significant difference [$F(1,11) = 2.466$, $p = 0.12$].

When comparing the result of Chinese subjects (Fig. 1a and b) to that of English subjects (Fig. 1c and d), both frequency ranges show that Chinese subjects (Fig. 1a and b) performed better overall than the English subjects in both UP and DOWN FM identification. For example, when FM duration decreased to 20 ms, the performance of English subjects dropped to around 80% correct or below, whereas the Chinese listeners remained at an identification rate of 98%. Chinese subjects were also more accurate than English subjects in DOWN FM identification as shown in Fig. 1 (dotted grey line). The DOWN FM performance of English subjects decreased below 80% at FM durations of 80 ms, whereas the 80% threshold in Chinese subjects was 5 ms. Regarding the direction performance asymmetry, Chinese subjects showed a reversed pattern compared to that of English subjects: English subjects detected UP FM better than DOWN FM, whereas Chinese subjects showed better performance for DOWN FM identification. Importantly, the FM duration/rate *region* in which the UP/DOWN asymmetry occurred differed between these two subject groups. As the FM duration decreased from 640 ms – where both groups performed well – the asymmetry appeared earlier for English subjects (around 80 ms) compared to Chinese subjects (below 20 ms).

Fig. 2 illustrates the *d*-prime and $\beta_{\mathrm{normalized}}$ results for the two language groups and for both the low (600–900 Hz) and high (2–3 kHz) ranges. As shown in Fig. 2a and b, for both frequency ranges, Chinese and English subjects both showed larger *d*-prime values for longer FM durations, which is consistent with percent correct performance in Fig. 1; larger *d*-prime values represent better FM direction identification ability. Importantly, Fig. 2a and b illustrates clearly that Chinese subjects showed significantly larger *d*-primes than English subjects for both low and high frequency ranges, especially at FM durations

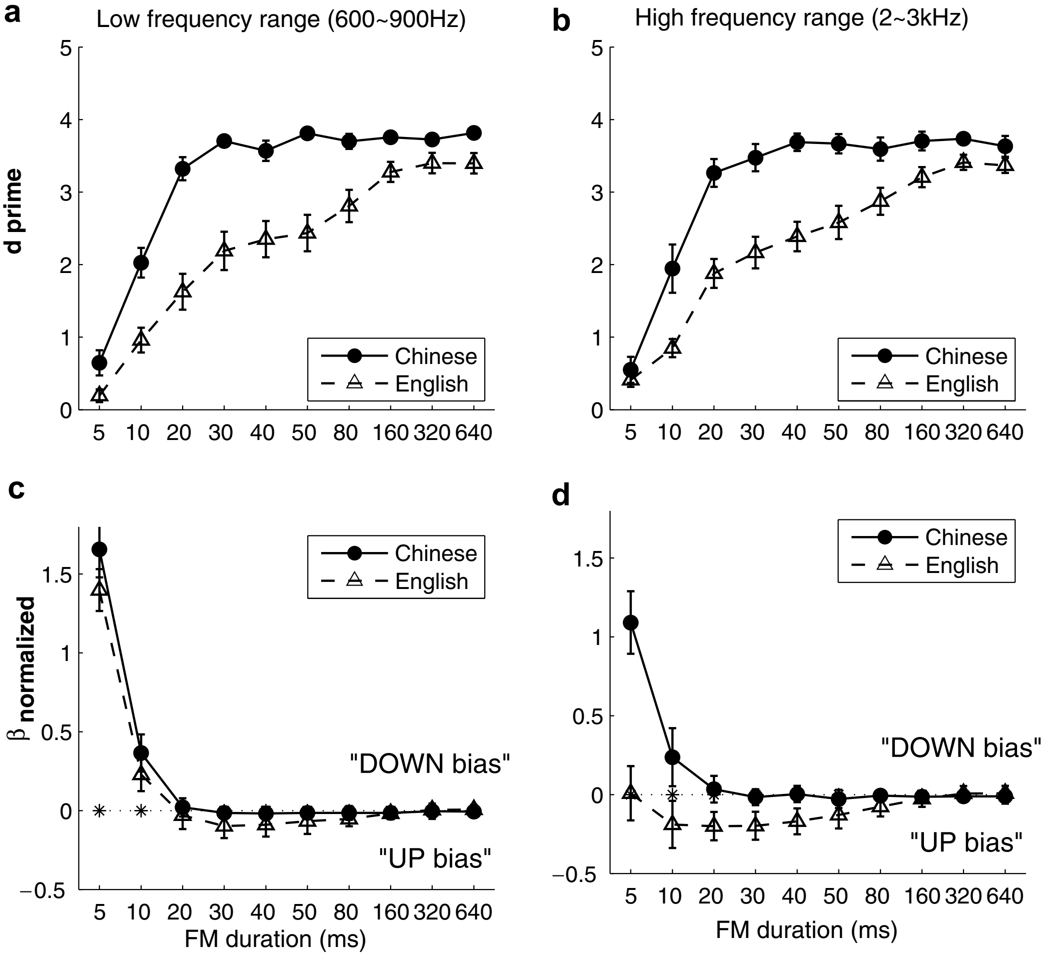

Fig. 2. d' (a, b) and $\beta_{\mathrm{normalized}}$ (c, d) results from Experiment 1 for both Chinese (solid line) and English (dotted line) subjects for direction identification of FM stimulus with different durations. The results for the low-frequency range (600–900 Hz) and high frequency range (2–3 kHz) are illustrated, respectively, in Fig. 2a and c and 2b and d. For the bias analysis (c, d), the starred line at zero represents the non-biased value, the area above it represents a 'DOWN bias', and area below it an 'UP bias'. Error bars are SEM.

from 10 ms to 160 ms, supporting the data pattern observed in Fig. 1. In summary, above FM durations of 160 ms, both subject groups performed similarly and similarly well; at the shortest tested FM durations of 5 ms, both groups performed similarly poorly. However, in the intermediate range of FM durations, values commensurate with acoustic variation of the speech signal, Chinese subjects manifested significantly better FM identification ability than English subjects.

The bias parameter, $\beta_{\mathrm{normalized}}$, illustrated in Fig. 2c and d, helps disentangle possible bias strategies from genuine responses, and accounts for some aspects of the results in Fig. 1, especially the UP versus DOWN direction asymmetry results. For example, as shown in Fig. 1a and b, there is a strong DOWN-better-than-UP pattern at the shortest FM durations (5 ms, 10 ms) for Chinese subjects, for both of the tested frequency ranges. Note that at FM duration of 5 ms, the UP FM percent correct in Fig. 1a and b actually dropped below chance level, indicating the possible involvement of a bias strategy in this condition. This is supported by the $\beta_{\mathrm{normalized}}$ values which are significantly above zero for FM durations of 5 ms and 10 ms (Fig. 2c and d) illustrating that Chinese subjects used a ‘press-DOWN’ strategy under sufficient stimulus uncertainty, i.e., when they in fact could not identify the fast FM direction. We suggest that such a bias results in the asymmetrical direction performance shown in Fig. 1a and b. It is also commensurate with the *d*-prime results in Fig. 2a and b which showed lower *d*-prime values at these two durations. As for English subjects, they showed below-zero $\beta_{\mathrm{normalized}}$ values during intermediate FM durations, indicating a ‘press-UP’ strategy for these FM signals – although their bias was smaller compared to Chinese subjects ($\mathrm{abs}(\beta_{\mathrm{normalized}}) < 0.5$). Therefore, the bias assumption can account for some part of the UP-better-than-DOWN data pattern for English subjects in Fig. 1c and d.

3.2. Experiment 2 (3IFC)

Fig. 3 shows the FM discrimination results for both Chinese (solid line) and English (dotted line) subjects. FM discrimination performance of the two subject groups overlapped completely. Analysis of variance showed an effect of duration [$F(8,40) = 53.7$, $p < 0.001$], but not an effect of native language [$F(1,5) = 1.23$, $p = 0.32$], and no significant interactions [$F(8,40) = 0.39$, $p = 0.92$]. Most importantly, both subject groups showed ceiling performance down to 20 ms; below 20 ms performance accuracy dropped off, although it remained above chance (>60% correct).

3.3. Frequency control

In Experiment 1, FM rate was varied by manipulating the duration of the stimuli, and subjects were asked to pay attention only to FM modulation direction in making their perceptual decision. However, subjects may have used the starting frequency as a cue in determining FM direction given that all the UP FM stimuli had lower starting frequencies than the DOWN FM stimuli. A control experiment was run to eliminate the possible frequency cues. No significant difference in results between this block and their corresponding results in the original experiments was observed indicating that listeners did not utilize starting frequency as a cue.

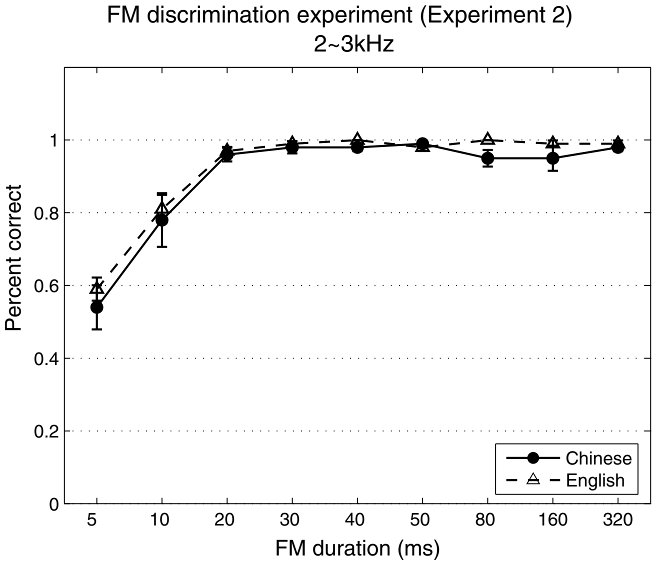

Fig. 3. Performance results for Mandarin Chinese (solid line) and English listeners (dotted line) on the three-interval-two-alternative-forced-choice (3IFC) paradigm. On each trial, participants heard three intervals, each of them containing FM sweeps of equal duration. Subjects indicated by button-press whether the FM in the second or third interval (forced-choice) was identical to the standard contained in the first interval. Error bars are SE.

4. Discussion

4.1. Multiple temporal scales of FM in speech

Spoken Chinese and spoken English incorporate FM acoustic information on at least two relevant temporal scales, namely FM at the segmental level (e.g., distinguishing [ga] from [pa] by assessing differences in formant transitions) and FM at the level of phrasal and sentential prosody (e.g., analysis of statement versus question in linguistic prosody, or analysis of happy versus angry intonation contours in affective prosody). However, these languages differ along another acoustic-phonetic dimension. Chinese (here Mandarin Chinese), as a tone language, incorporates tonal contrasts in its lexical phonology. Low-frequency modulation patterns (~80–400 Hz) of the fundamental frequency F0 form the phonological basis for lexical contrastiveness, a feature not observed in non-tone languages, for example English. In other words, Chinese and English listeners differ in that lexical items have different meanings as a function of their tonal contour in Chinese but not in English, and the FM contours germane to lexical tone are typically at longer (syllabic, ∼150–300 ms duration) time scales than the FM relevant to short formant transitions in stop consonants (∼15–50 ms).

Given this cross-linguistic state of affairs, a natural psycholinguistic and psychoacoustic question to ask is whether extensive processing expertise at one timescale can transfer to perceptual analysis at another time scale. Are tone-language speakers/listeners 'better' at hearing other auditory signals as a function of native language? In a non-speech FM direction identification experiment, we observed similar ceiling performance at FMs with longer durations (∼640 ms), in terms of both percent correct and *d*-prime results, across speakers of different languages. However, as FM duration decreased and FM signals became increasingly fast and more difficult to track, Chinese listeners showed a higher sensitivity to FM direction identification. One simple hypothesis is that the expertise of Chinese listeners in tracking FM at longer (syllabic and lexical) time scales – because of their tone language experience – was transferred to their performance at decoding FM in non-speech tasks, even though the time scales are different. One possible interpretation is the extensive ecologically natural experience with FM modulation sensitizes FM neurons in auditory cortex to respond to a wider time scale than is characteristic of non-tone-language speakers.

4.2. Specific low-level acoustic task shaped by native language experience

Based on the results from Experiment 1, we propose that the relevance of FM for Chinese speakers/listeners in decoding every lexical item may have as a consequence altered (lowered) thresholds in the perceptual analysis of relatively basic psychoacoustic tasks with FM tones or glides. Importantly, FM direction identification tasks are not ecologically natural perceptual tasks in human audition, especially for non-tone-language subjects, because when listeners hear these kinds of FM sounds, they do not explicitly name the direction of such sounds (UP or DOWN) if not required – although they may be aware of the acoustic differences. We hypothesize that a possible reason for the better performance of Chinese subjects than that of English subjects in Experiment 1 lies in their language experience in explicit FM direction naming rather than their better basic temporal processing ability. In other words, we assume that the basic temporal processing thresholds are the same across subjects because they are simply fundamental architectural properties of the auditory system, but that task demands in auditory studies can tap into experience-dependent differences among subject groups.

The identical performance between these two language groups in Experiment 2, the FM discrimination experiment, supports this alternative explanation. In this experiment, subjects were not required to explicitly name FM stimulus direction (as was required in Experiment 1), but only had to discriminate between two sounds. In order to accomplish such a discrimination, listeners can employ a large number of acoustic cues in the trial to make the judgment whether the two sounds are exactly the same or different, and therefore Chinese subjects lose the advantage of their expertise in naming FM directions. And, as we observed, in this sensitive 3IFC task, the thresholds between the two subject groups were the same. In summary, we conjecture that Chinese subjects performed better in the FM identification task because they are experts in explicit FM direction *labeling* by being tone-language speakers, and it is a task-dependent language advantage rather than a true fundamental temporal processing advantage that accounts for the FM identification performance difference between these two language groups. Our findings are also consistent with many cross-linguistic neuroimaging experiments investigating whether specific language experience can alter the deployment of neural circuits for auditory signal processing (Gandour et al., 1998, 2000, 2002; Scott, 2004; Hsieh et al., 2001; Klein et al., 2001; Lee and Nusbaum, 1993; Wong et al., 2004; Xu et al., 2006). For example, an fMRI study (Gandour et al., 2002) comparing Thai and English subjects in a duration judgment task under both speech and non-speech context showed that encoding of complex auditory signals is influenced by their functional role in a particular language, and specifically Thai vowel length in that case. It has been shown in previous physiological studies that Chinese subjects exhibited stronger pitch representation and smoother pitch tracking than the English subjects at brain stem level (Krishnan et al., 2004, 2005), suggesting that the language experience may enhance processing of linguistically relevant features of input.

4.3. 20 ms as a common temporal processing threshold

The 20 ms duration boundary observed very sharply in Experiment 2 replicates work by others concerning FM direction detection (Schouten, 1985, 1986, 1989). Furthermore, this number is near the 'order threshold' to distinguish the temporal order between two clicks (Hirsh, 1959) and has also been documented as relevant for many other psychophysical and neurophysiological phenomena (Eggermont, 1998; Kanabus et al., 2002; Miller et al., 2004; Michalewski et al., 2005; Zampini et al., 2005), leading us to favor the view that processing at the 20 ms time scale reflects the construction of elementary auditory percepts (Poeppel, 2003). A potential counterexample to this view, a study reporting that duration is not relevant, by Madden and Fire (1997), tested only two durations and was thus not able to investigate in a parametric manner to what extent a duration manipulation is actually relevant.

A fundamental question concerns whether the results we report on the importance of a 20 ms threshold reflect a temporal processing property (or *temporal integration constant*) or a rate processing property (which is a combinatorial characteristic of the frequency and time domains). In other words, does the 'threshold' (20 ms or 25 oct/s) provide an index of people's *temporal* processing ability or their *FM rate sensitivity*? Our preferred interpretation is that these data reflect a primarily temporal processing ability. Of course, in our study as in all others, duration and FM rate cannot be disentangled in principle. We chose to parameterize rate as duration because we wanted to keep the frequency span within values typical for speech formants – e.g., typical F1 range, 600–900 Hz. The alternative, to hold stimulus duration constant and vary FM rate by changing the frequency span, would introduce frequency modulations entirely atypical for speech sounds and therefore not of primary interest in this study. In light of this perspective, half an octave frequency span is large enough for subjects to detect FM at most stimulus durations. When duration is very short (below 20 ms), even if one increases the frequency span, FM detection ability does not change dramatically. This can be seen in Schouten's result when duration is 15 ms, the "Down preference" disappears, and the Down response percent was always nearly 50% no matter the actual direction and FM rate (Schouten and Pols, 1989). Therefore, we take 20 ms to reflect a *fundamental temporal integration window for the construction of auditory percepts*, regardless of whether listeners are tone-language or non-tone-language speakers. We hypothesize (i) that it is a general threshold which limits our ability to process auditory stimuli, and (ii) that it is possibly a hardwired architectural property of our central auditory system and therefore cannot be altered by environmental factors (Poeppel, 2003).

4.4. FM direction selectivity and 'bias strategy'

One final issue to consider concerns to what extent the present data converge with or differ from previous data of the same type, for example the data reported by Gordon and Poeppel (2002), where the same materials and procedure (of Experiment 1) were used for English speakers (see Fig. 1). In the earlier report, the ability to detect UP was much better than DOWN at stimulus durations above 20 ms, and we interpreted the up/down asymmetry finding, in particular, as the result of a low-level (basilar membrane) explanation for such a behavioral discrimination performance, consistent with the position espoused by Collins and Cullen (1978) and Dau et al. (2000). However, such FM direction selectivity is always confused and affected by possible subject bias strategy, especially in a binary judgment task (e.g., 2AFC here), and by calculating two important parameters *d*-prime and $\beta_{\text{normalized}}$ in signal detection theory, we can get a clearer understanding of the factors underlying performance. For example, from *d*-prime and $\beta_{\text{normalized}}$ we can infer that the dramatic DOWN-better-than-UP performance for 5 ms FM in Chinese subjects is due to their dramatic bias to press "DOWN" (Fig. 2c and d) under that condition. The real direction of UP/DOWN asymmetry has appeared in many other experiments – sometimes in the opposite direction (Dooley and Moore, 1988; Schouten, 1985, 1986) – and the reasons for such an UP/DOWN asymmetry require further study.

5. Conclusions

Motivated by questions on experience-dependent plasticity in speech perception and hearing as a function of native language, we performed a cross-linguistic non-speech FM perception experiment. We administered two psychophysical tasks (single-trial 2AFC and 3IFC), testing the identification and discrimination of FM sweeps in Chinese (here Mandarin Chinese) and English listeners. Chinese and English participants were selected because of their differential use of FM acoustic information in their native languages. Our results demonstrate (i) that Chinese subjects are better than English subjects in FM direction identification, a task requiring explicit FM direction naming which is frequently involved in tone-language environment, but (ii) that Chinese and English listeners were the same in an FM discrimination task, a more elementary auditory temporal process similar for different language groups. In summary, our data confirmed task-dependent, language-specific effects on the perceptual processing of simple acoustic FM stimuli and support the hypothesis of a relatively common basic temporal processing threshold ($\sim$20 ms) regardless of native language.

Acknowledgements

This work is supported by NIH R01 DC05660 to DP, and we are grateful to two reviewers for thoughtful and detailed comments.

References

Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat. Neurosci. 8 (3), 389–395.

Collins, M.J., Cullen, J.K.J., 1978. Temporal integration of tone glides. J. Acoust. Soc. Am. 63 (2), 469–473.

Dau, T., Wegner, O., Mellert, V., Kollmeier, B., 2000. Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion. J. Acoustic. Soc. Am. 107 (3), 1530–1540.

Dimitrijevic, A., Sasha John, M., van Roon, P., Picton, T.W., 2001. Human auditory steady-state responses to tones independently modulated in both frequency and amplitude. Ear Hear. 22 (2), 100–111.

Dooley, G.J., Moore, B.C., 1988. Duration discrimination of steady and gliding tones: a new method for estimating sensitivity to rate of change. J. Acoust. Soc. Am. 84 (4), 1332–1337.

Dorfman, D.D., Alf, E.J., 1968. Maximum likelihood estimation of parameters of signal detection theory – a direct solution. Psychometrika 33 (1), 117–124.

Doupe, A.J., Kuhl, P.K., 1999. Birdsong and human speech: common themes and mechanisms. Annu. Rev. Neurosci. 22, 567–631.

Edwards, B.W., Viemeister, N.F., 1994a. Frequency modulation versus amplitude modulation discrimination: evidence for a second frequency modulation encoding mechanism. J. Acoust. Soc. Am. 96 (2 Pt 1), 733–740.

Edwards, B.W., Viemeister, N.F., 1994b. Psychoacoustic equivalence of frequency modulation and quasi-frequency modulation. J. Acoust. Soc. Am. 95 (3), 1510–1513.

Edwards, B.W., Viemeister, N.F., 1997. Masking of a brief probe by sinusoidal frequency modulation. J. Acoust. Soc. Am. 101 (2), 1010–1018.

Eggermont, J.J., 1994. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. Hear. Res. 74 (1–2), 51–66.

Eggermont, J.J., 1998. Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J. Neurophysiol. 80 (5), 2743–2764.

Eggermont, J.J., 2001. Between sound and perception: reviewing the search for a neural code. Hear. Res. 157 (1–2), 1–42.

Gandour, J., Wong, D., Hutchins, G., 1998. Pitch processing in the human brain is influenced by language experience. Neuroreport 9 (9), 2115–2119.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., Hutchins, G.D., 2000. Ib crosslinguistic PET study of tone perception. J. Cognitive Neurosci. 12 (1), 207–222.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Long, Y.X., et al., 2002. Neural circuitry underlying perception of duration depends on language experience. Brain Lang. 83 (2), 268–290.

Gordon, M., Poeppel, D., 2002. Inequality in identification of direction of frequency change (up vs. down) for rapid frequency modulated sweeps. Acoust. Res. Lett. Online 3 (1), 29–34.

Heil, P., Rajan, R., Irvine, D.R., 1992. Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I: Effects of variation of stimulus parameters. Hear. Res. 63 (1–2), 108–134.

Hirsh, I.J., 1959. Auditory perception of temporal order. J. Acoustic. Soc. Am. 31 (6), 759–767.

Howie, J.M., 1976. Acoustical studies of Mandarin Vowels and Tones. University Press, Cambridge.

Hsieh, L., Gandour, J., Wong, D., Hutchins, G.D., 2001. Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang. 76 (3), 227–252.

John, M.S., Dimitrijevic, A., van Roon, P., Picton, T.W., 2001. Multiple auditory steady-state responses to AM and FM stimuli. Audiol. Neurootol. 6 (1), 12–27.

John, M.S., Dimitrijevic, A., Picton, T.W., 2002. Auditory steady-state responses to exponential modulation envelopes. Ear Hear. 23 (2), 106–117.

Kanabus, M., Szelag, E., Rojek, E., Poppel, E., 2002. Temporal order judgement for auditory and visual stimuli. Acta Neurobiol. Exp. 62 (4), 263–270.

Klein, D., Zatorre, R.J., Milner, B., Zhao, V., 2001. A cross-linguistic PET study of tone perception in mandarin Chinese and English speakers. Neuroimage 13 (4), 646–653.

Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P.A., 2004. Human frequency-following response: representation of pitch contours in Chinese tones. Hear. Res. 189 (1–2), 1–12.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. 25 (1), 161–168.

Lee, L., Nusbaum, H.C., 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. Percept. Psychophys. 53 (2), 157–165.

Liang, L., Lu, T., Wang, X., 2002. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J. Neurophysiol. 87 (5), 2237–2261.

Liberman, A.M., Delattre, P.C., Gerstman, L.J., Cooper, F.S., 1956. Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. Exp. Psychol. 52 (2), 127–137.

Luo, H., Wang, Y., Poeppel, D., Simon, J.Z., 2006. Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. J. Neurophysiol.

Madden, J.P., Fire, K.M., 1997. Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency. J. Acoust. Soc. Am. 102 (5 Pt 1), 2920–2924.

Maiste, A., Picton, T., 1989. Human auditory evoked-potentials to frequency-modulated tones. Ear Hear. 10 (3), 153–160.

Makela, J.P., Hari, R., Linnankivi, A., 1987. Different analysis of frequency and amplitude modulations of a continuous tone in the human auditory cortex: a neuromagnetic study. Hear. Res. 27 (3), 257–264.

Mendelson, J.R., Schreiner, C.E., Sutter, M.L., Grasse, K.L., 1993. Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. Exp. Brain Res. 94 (1), 65–87.

Michalewski, H.J., Starr, A., Nguyen, T.T., Kong, Y.Y., Zeng, F.G., 2005. Auditory temporal processes in normal-hearing individuals and in patients with auditory neuropathy. Clin. Neurophysiol. 116 (3), 669–680.

Miller, J., Kuhlwein, E., Ulrich, R., 2004. Effects of redundant visual stimuli on temporal order judgments. Percept. Psychophys. 66 (4), 563–573.

Moore, B.C., 1997. An Introduction of Psychology of Hearing, fourth ed. Academic press.

Moore, B.C., Sek, A., 1992. Detection of combined frequency and amplitude modulation. J. Acoust. Soc. Am. 92 (6), 3119–3131.

Picton, T.W., Skinner, C.R., Champagne, S.C., Kellett, A.J., Maiste, A.C., 1987. Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. J. Acoust. Soc. Am. 82 (1), 165–178.

Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D., 2003. Human auditory steady-state responses. Int. J. Audiol. 42 (4), 177–219.

Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun. 41 (1), 245–255.

Rosenblith, W.A., Stevens, K.N., 1953. On the DL for frequency. The J. Acoustic. Soc. Am. 25 (9), 980–985.

Rupp, A., Uppenkamp, S., Gutschalk, A., Beucker, R., Patterson, R.D., Dau, T., et al., 2002. The representation of peripheral neural activity in the middle-latency evoked field of primary auditory cortex in humans. Hear. Res. 174 (1–2), 19–31.

Schouten, M.E., 1985. Identification and discrimination of sweep tones. Percept. Psychophys. 37 (4), 369–376.

Schouten, M.E., 1986. Three-way identification of sweep tones. Percept. Psychophys. 40 (5), 359–361.

Schouten, M.E., Pols, L.C., 1989. Identification and discrimination of sweep formants. Percept. Psychophys. 46 (3), 235–244.

Scott, S., 2004. Does language affect the neural basis of speech perception? – A comparison of English and Mandarin speakers. Int. J. Psychol. 39 (5–6), 387.

Stagray, J.R., Downs, D., Sommers, R.K., 1992. Contributions of the fundamental, resolved harmonics, and unresolved harmonics in tone-phoneme identification. J. Speech Hear. Res. 35 (6), 1406–1409.

Suga, N., 1968. Analysis of frequency-modulated and complex sounds by single auditory neurones of bats. J. Physiol. 198 (1), 51–80.

Swets, J., 1982. Recent theoretical developments in signal-detection and recognition. Psychophysiology 19 (3), 300.

Whitfield, I.C., Evans, E.F., 1965. Responses of auditory cortical neurons to stimuli of changing frequency. J. Neurophysiol. 28, 655–672.

Wong, P.C., Parsons, L.M., Martinez, M., Diehl, R.L., 2004. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. J. Neurosci. 24 (41), 9153–9160.

Xu, Y.S., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y.X., et al., 2006. Activation of the left planum temporale in pitch processing is shaped by language experience. Human Brain Mapp. 27 (2), 173–183.

Zampini, M., Brown, T., Shore, D.I., Maravita, A., Roder, B., Spence, C., 2005. Audiotactile temporal order judgments. Acta Psychol. 118 (3), 277–291.