Pitch normalization in speech and music

Kent Tokunaga, Corey E. Young, Lori Ezzedine,
Ekaterina Akimenko, Hope E. Schmid & Evan D. Bradley

Background

Pitch in Language and Music

- Language and music both use pitch[1], for lexical tones (Figure 1) and intervals (Figure 2).
- Lexical tone is unfamiliar and hard to learn for speakers of non-tonal languages (English).[1]
- Musicians are better than others at many pitch tasks, including lexical tone perception.[2]

Fig 1. Mandarin[3]
1) ma1 'mother'
2) ma2 'hemp'
3) ma3 'horse'
4) ma4 'scold'

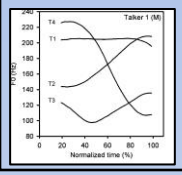
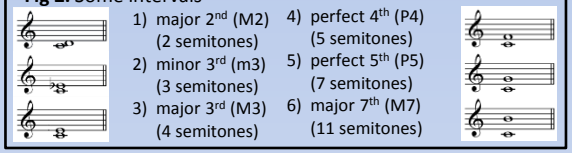


Fig 2. Some intervals



1) major 2nd (M2) (2 semitones)
2) minor 3rd (m3) (3 semitones)
3) major 3rd (M3) (4 semitones)
4) perfect 4th (P4) (5 semitones)
5) perfect 5th (P5) (7 semitones)
6) major 7th (M7) (11 semitones)

Speaker Normalization

- Phonetic variation between and within talkers means even the same word can sound rather different, depending on context.[4,5]
- Listeners need to compensate for (normalize) this variability, and normalization ability improves with language proficiency.[4,5]
- Stimulus variability plays a role in the learning of tones.[6]
  - English speakers with better pitch perception learned Mandarin tones best with high variability training sets.[6]
  - English speakers with poorer pitch perception learned best with low variability training sets.[6]
- Learners with poorer pitch perception may not yet perceive pitch well enough to recognize tones spoken by different speakers.[3]
- Music perception requires normalization processes similar to those found in language.[7]
  - Musical elements maintain their musical identity across different instruments or keys.[7]

Hypotheses

We explored the ability of participants to perceive lexical tones and musical intervals at different levels of variability.

1. We expected that, for non-native listeners, lexical tone perception ability decreases with increasing variability.
2. If music is subject to similar demands, interval perception should also decrease with greater variability.

Experiment 1 (AX)

Tone Discrimination Test

Words: monosyllables ('ka', 'ma', 'di') spoken with four tones (Figure 1)
- 4 Native Mandarin speakers from China (3 female, 1 male)
- Some were real words, some not

AX (discrimination) Trials:
- 2 words with same syllable
- Same (40%) or different (60%) tone
- 1500ms inter-onset interval

Blocks: 4 variability levels:
- **1:** same speaker
- **2:** two speakers , same gender
- **2.5:** two speakers, different genders
- **3:** four speakers (3 female, 1 male)

Interval Discrimination Test

Intervals: melodic M2, m3, M3, P4 (Fig. 2)
- Piano or saxophone timbre
- Ascending & descending directions
- Played in one of two adjacent octaves
- Synthesized using *Aria Maestosa*

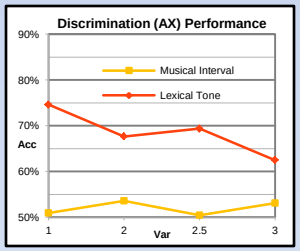
AX (discrimination) Trials:
- Same direction, different starting pitch
- Same (40%) or different (60%) interval
- 1000ms inter-onset interval

Blocks: 4 variability levels:
- **1:** same instrument & octave
- **2:** both instruments, same octave
- **2.5:** same instrument, both octaves
- **3:** both instruments & octaves

- 12 Blocks in each test (3 at each variability level) of 10 trials each (120 trials).
- Participants heard examples of the four tones /intervals, and an explanation of
  a) the importance of pitch in determining word meaning in this language.
  b) the concept of size (distance between notes) as the key to interval identity.
- Participants decided whether the words/intervals were the "same" or "different".
- Both tests administered to 30 Penn State students.
  - Within a single 30-minute session; counterbalanced for order.
  - Audio presented via speakers or headphones; responses made on paper form.
- Participants completed language and music history questionnaires.
  - Primarily monolingual English-speaking ; some other non-tonal languages.
  - Overall low degree of musicianship.

Results

- As variability increased, tone discrimination decreased as predicted.
  - No difference due to gender at Level 2.
- Interval discrimination was not affected.
  - Near chance for all variability levels.



Separate mixed-effects ANOVAs
accuracy ~ variability | subject
tone F(1,28)=18.813, p<.001
interval F=(1,28)=0.1669, p>.6

Why can't intervals be discriminated?

Intervals should be more familiar than tones!

a) Intervals are too similar in frequency ratio?
  - Frequency differences are smaller than those found in many tone systems.[8]

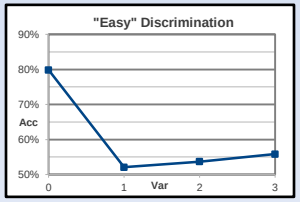
b) Task is too different from word perception?
  - "Same/different word" may be more familiar/intuitive than "interval size".

"Easy" Interval Discrimination Test (N=35)

Intervals: melodic M2, P4, P5, M7 (Fig. 2)

Blocks: less variability; one direction per block
- **0:** same start note *(not true interval test!)*
- **1-3:** levels 1, 2, 3 from above

- Still near chance for all true interval levels.
- Experiment 2 tests task effects on same stimuli



Experiment 2 (XAB)

Tone Matching Test

Words: same as Experiment 1

XAB (match-to-sample) Trials:
- 3 words with same syllable
- 2500ms inter-onset interval X to A
- 1500ms inter-onset interval A to B

Blocks: 2 variability levels:
- **LO:** X, A, & B had same speaker; 1 speaker per block
- **HI:** X one speaker, A & B another; 4 speakers per block

Interval Matching Test

Intervals: ascending M2, P4, P5, M7

XAB (match-to-sample) Trials:
- X started on one note, A & B another.
- 2500ms inter-onset interval A to X
- 1500ms inter-onset interval A to B

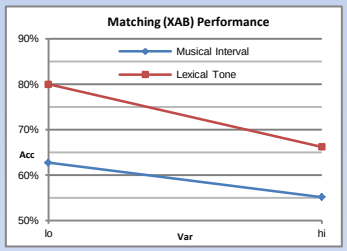
Blocks: 2 variability levels:
- **LO:** X in same instrument and same octave as A & B
- **HI:** X in a different instrument and octave than A & B

- 8 Blocks in each test (4 at each variability level) of 12 trials each (84 trials).
- Participants heard same examples and explanation as Experiment 1.
  - Interval test included a "warmup" block with all intervals on the same note.
- Participants decided which word/interval (A/B) was the same as the standard (X).
- Tests administered in same manner as Experiment 1 to 23 Penn State students.
  - Music and language demographics similar to Experiment 1.

Results

- As stimulus variability increased, tone matching performance decreased.
- As stimulus variability increased, interval matching performance decreased.

Separate mixed-effects ANOVAs
accuracy ~ variability | subject
tone F(1,21)=27.436, p<.001
interval F=(1,21)=19.108, p<.001



Summary

- We tested whether the ability to adapt to stimulus variability is correlated across novel musical and linguistic pitch tasks for English-speaking nonmusicians.
- Hearing more speakers decreased perception of foreign lexical tones in both discrimination (AX) and matching (XAB) tasks, consistent with predictions.
- Melodic intervals could only be reliably perceived in a matching task.

How similar is normalization for linguistic and musical pitch? Hypotheses:

a) Pitch is normalized differently in linguistic vs. musical tasks (by English speakers);
  - This could be due to different relevant cues for tones/intervals.

b) Discrimination may encourage different processing of words vs. intervals;
  - Even familiar intervals may be hard for nonmusicians outside of melodies.

- Future studies will compare participants by musicality and language to determine how linguistic and musical experience affects normalization.

1. Bradley (2013). *RRHLM* 1.
2. Patel (2011). *FPSYG 2*(142).
3. Chandrasekaran (2010). *JASA 128*(1).
4. Johnson (2005).
5. Xu (1997). *J Phon 25*(1).
6. Lee (2007). *ICPhS*.
7. Tervaniemi (2001). *Learn Memory 8*(5).
8. Maddieson (1991). *YPL* 15.