Perceiving Prosody in Speech

Effects of Music Lessons

WILLIAM FORDE THOMPSON, E. GLENN SCHELLENBERG, AND GABRIELA HUSAIN

Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario L5L 1C6, Canada

ABSTRACT: In two experiments, musically trained and untrained adults were tested on their ability to match spoken utterances with their tonal analogues (tone sequences that retained the pitch and temporal patterns of the utterances). In both cases, musical training was associated with superior performance, indicating an enhanced ability to extract prosodic information from spoken phrases.

KEYWORDS: speech; prosody; music lessons

INTRODUCTION

We conducted two experiments that tested whether formal training in music is associated with an enhanced ability to perceive prosody in speech. In other words, we examined the possibility that music lessons have positive transfer effects that influence speech perception. Relatively few studies have examined the effects of music lessons on nonmusical auditory skills. Nonetheless, training in music undoubtedly refines and develops mechanisms involved in processing music. Such mechanisms could extend to other complex auditory signals that are characterized by variation in dimensions relevant to music. In particular, prosody in speech involves the use of paralinguistic cues (e.g., pitch variation, speed, and amplitude changes) that serve to highlight the syntax and the semantics of an utterance as well as the speaker's emotional state. Paralinguistic cues, such as pitch height, speed of talk, and loudness, are essential for accuracy in decoding emotion in speech.[1,2] These cues are also known to be important in conveying emotion through music.[3]

Decoding speech prosody involves an initial stage in which prosodic information is extracted and represented and a subsequent stage in which meaning is attributed to that information. The experiments reported here examined whether music lessons yield an enhanced ability to *extract* prosodic cues in speech.

Address for correspondence: Dr. William Forde Thompson, Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario L5L 1C6, Canada. Voice: 905-569-4733; fax: 905-569-4734.

b.thompson@utoronto.ca

Ann. N.Y. Acad. Sci. 999: 530–532 (2003). © 2003 New York Academy of Sciences.
doi: 10.1196/annals.1284.067

METHODS AND RESULTS

In Experiment 1, musically trained ($n = 22$) and untrained ($n = 16$) listeners were tested on their ability to extract pitch information from short phrases spoken in English (e.g., "the boy went to the store"). On average, trained listeners had more than 15 years of music lessons. Stimuli included four semantically neutral but happy-sounding utterances taken from the Florida Affect Battery (FAB) as well as intonation melodies (tone sequences) derived from these utterances. Each intonation melody consisted of tones that matched the modal pitch and duration of the spoken syllables. Tone syllables had equal amplitude and their pitch varied discretely (unlike continuous pitch variations in speech). In short, the intonation melodies were an abstraction of the pitch and temporal variations in the utterances. "Incorrect" melodies that did not match the utterances were formed by switching the location of two tone syllables. On each trial, listeners heard a spoken phrase followed by an intonation melody. They judged whether or not the intonation melody matched the prosody of the phrase. Musically trained participants outperformed their untrained counterparts, $t(36) = 2.47$, $P < 0.02$, with both groups exceeding chance levels of performance.

Experiment 2 was similar to the first except that: (1) the speech stimuli were taken from a foreign language (Spanish), (2) the paralinguistic cues of the speech stimuli conveyed a happy or a sad emotion, and (3) the intonation melodies were created by means of high-pass filtering, such that the pitch of the intonation melodies varied continuously rather than discretely. We tested 12 musically trained ($M = 16$ years of lessons) and 8 musically untrained listeners, none of whom spoke Spanish. Mismatching intonation melodies were derived from novel sentences spoken in the same emotional tone as the target sentences. There was no main effect of emotion (happy and sad) and no interaction between emotion and training group. As in Experiment 1, the musically trained group performed better than the untrained group, $F(1,18) = 4.76$, $P < 0.05$, with both groups exceeding chance levels of responding.

DISCUSSION

In both experiments, musically trained participants were better than untrained participants at extracting prosodic information from speech. This enhanced ability to extract prosodic information extended to phrases spoken in an unfamiliar language, and the effect was evident regardless of whether the pitch information was represented discretely or continuously. Such evidence of cognitive transfer between music and speech raises the possibility that speech prosody and music are processed with shared neural resources.[4]

Other evidence suggests that music lessons not only improve the ability to *extract* prosodic cues, but also improve the ability to *interpret* speech prosody. In one study,[5] music and law students listened to auditory signals comprised of the fundamental frequencies of voice samples recorded from depressed and nondepressed individuals (i.e., the words were inaudible). Music students were better than law students at identifying the emotional state of the speakers. More recently,[6] we confirmed that children and adults with music training have enhanced skill at interpreting the emotional significance of prosodic information. Musical expertise requires the ability to express emotional information through variations in loudness, timing, and pitch, and this ability appears to be reflected in an enhanced skill at *extracting* and *interpreting* emotional meaning in speech.

ACKNOWLEDGMENTS

This research was supported by a grant (to W.F.T.) from the International Foundation for Music Research. We thank Cory Sand and Don Smith for research assistance.

REFERENCES

1. FRICK, R. 1985. Communicating emotion: the role of prosodic features. Psychol. Bull. **97:** 412–429.
2. PLANALP, S. 1998. Communicating emotion in everyday life: cues, channels, and processes. *In* Handbook of Communication and Emotion: Research, Theory, Applications, and Contexts. P.A. Andersen & L.K. Guerrero, Eds. :29–48. Academic Press, Inc. San Diego, CA.
3. JUSLIN, P.N. & J.A. SLOBODA. 2000. The influence of musical structure on emotion expression. *In* Music and Emotion. Theory and Research. P.N. Juslin & J.A. Sloboda, Eds. :223–248. Oxford University Press. New York, NY.
4. PATEL, A.D., I. PERETZ, M. TRAMO & R. LABRECQUE. 1998. Processing prosodic and music patterns: a neuropsychological investigation. Brain & Lang. **61:** 123–144.
5. NILSONNE, A. & J. SUNDBERG. 1985. Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. Music Percept. **2:** 507–516.
6. THOMPSON, W.F., E.G. SCHELLENBERG & G. HUSAIN. 2004. Decoding speech prosody: do music lessons help? Emotion **4**(1). In press.