Processing simultaneous auditory objects: Infants' ability to detect mistuning in harmonic complexes

Nicole A. Folland and Blake E. Butler
Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada

Nicholas A. Smith
Perceptual Development Laboratory, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

Laurel J. Trainor[a)]
Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada

(Received 22 December 2010; revised 27 April 2011; accepted 29 April 2011)

The ability to separate simultaneous auditory objects is crucial to infant auditory development. Music in particular relies on the ability to separate musical notes, chords, and melodic lines. Little research addresses how infants process simultaneous sounds. The present study used a conditioned head-turn procedure to examine whether 6-month-old infants are able to discriminate a complex tone (240 Hz, 500 ms, six harmonics in random phase with a 6 dB roll-off per octave) from a version with the third harmonic mistuned. Adults perceive such stimuli as containing two auditory objects, one with the pitch of the mistuned harmonic and the other with pitch corresponding to the fundamental of the complex tone. Adult thresholds were between 1% and 2% mistuning. Infants performed above chance levels for 8%, 6%, and 4% mistunings, with no significant difference between conditions. However, performance was not significantly different from chance for 2% mistuning and significantly worse for 2% compared to all larger mistunings. These results indicate that 6-month-old infants are sensitive to violations of harmonic structure and suggest that they are able to separate two simultaneously sounding objects. © *2012 Acoustical Society of America*.
[DOI: 10.1121/1.3651254]

PACS number(s): 43.75.Cd, 43.66.Hg [DD] Pages: 993–997

I. INTRODUCTION

Much music contains multiple simultaneous notes whether from different instruments playing simultaneously or one instrument playing several notes at the same time. However, the sound wave that reaches the ear is a composite of the sound energy from each source. Therefore, the auditory system must analyze the spectrotemporal properties of the input in order to determine the number and identity of the auditory objects that gave rise to it (Bregman, 1990). This process of parsing out distinct auditory events and sources to create meaningful percepts is known as *auditory scene analysis* (Bregman, 1990). One cue to whether a set of frequency components originated from one source is the harmonic relations between them. Sounds with pitch typically have energy at harmonics that are integer multiples of a fundamental. Reflecting this, the auditory system tends to integrate frequency components standing in integer relations into a single object or percept (Hartmann, 1996; Lin and Hartmann, 1998). Conversely, when a frequency component does not fit with a set of such components it will tend to be heard as a separate auditory object originating from a different source.

In music, analysis of frequency relations is critical for perceiving pitch, as well as for segregating the different musical sounds in polyphonic music, for the perception of harmony, and for perceiving consonance and dissonance. With respect to the last point, recent research indicates that perception of consonance is more highly correlated with whether the harmonics are integer multiples of a fundamental (McDermott *et al.*, 2010) than whether there is beating or roughness present, as suggested by Plomp and Levelt (1965). Thus, analysis of harmonicity appears to be crucial for pitch perception, for identification of simultaneous auditory objects, and for the perception of consonance and dissonance.

Frequency processing in the cochlea appears to be mature at birth (Teas *et al.*, 1982; Abdala and Chatterjee, 2003). Psychophysical and auditory brainstem studies of the frequency resolution of spatial encoding mechanisms concur that 3-month-old infants show adult-like processing for frequencies below 4000 Hz, and that by 6 months of age infants show mature processing for all frequencies (Abdala and Folsom, 1995; Folsom, 1985; Folsom and Wynne, 1987; Olsho, 1985; Spetner and Olsho, 1990). Pitch discrimination thresholds for high frequency pure tones (over 4000 Hz) also rely predominantly on spatial encoding mechanisms, and follow a similar developmental timeline as for frequency resolution (Olsho *et al.*, 1987). However, pitch discrimination at lower

[a)]Author to whom correspondence should be addressed. Electronic mail: ljt@mcmaster.ca

frequencies relies additionally on temporal mechanisms and remains immature until around 10 or 11 years of age (Maxon and Hochberg, 1982; Werner, 2007). Nonetheless, pitch discrimination at low frequencies is mature enough in early infancy to support the fine discrimination needed for music perception (Trainor and Corrigall, 2010; Trainor and He, 2011). At 2 months of age, infants are able to discriminate vowel sounds that differ in frequency by 3% (Swoboda *et al.*, 1976) and show cortical electroencephalograph (EEG) responses to a 6% change in the pitch of piano tones (He *et al.*, 2007). By 4 months infants show EEG responses to a 3% change in the pitch of guitar tones (Trainor *et al.*, 2011) and by 5 months can behaviorally discriminate a 2% change in the frequency of pure tones (Olsho *et al.*, 1982).

Whether infants integrate frequency components into a single pitch percept is a more difficult question. In adults, when the auditory system is presented with several harmonics of a complex tone in the absence of the fundamental frequency, the pitch of the complex sound can still be perceived even though there is no energy at the fundamental frequency. Using event-related potentials derived from EEG recordings, He *et al.* (2009) showed that by 4 months of age infants also perceive the pitch of the missing fundamental.

In the present study we examined infants' sensitivity to violations in harmonicity that lead to the perception of two auditory objects in adults. Previous behavioral studies with adults indicate an improved ability to segregate a mistuned harmonic with increased length of stimulus, increased amount of mistuning, and decreased harmonic number. Young adults (aged 22–24 years) have thresholds that vary between 0.5% and 8% mistuning depending on the stimuli used, while children (aged 8–13 years) and the elderly are less sensitive and report fewer instances of hearing two objects across all amounts of mistuning (Hartmann *et al.*, 1990; Alain *et al.*, 2001; Alain *et al.*, 2003; Alain and McDonald, 2007). EEG and MEG (magnetoencephalography) studies in adults and children (aged 8–13) have revealed a neural component that is associated with the perception of two auditory objects and is thus referred to as the object-related negativity (ORN). It has a frontally negative and posteriorly positive topography consistent with activation in auditory areas and is present regardless of stimulus probability (Alain and Schuler, 2002). The ORN occurs in the absence of attention or awareness so it is thought to reflect a bottom-up or low-level process that is largely unconscious, such that the mistuning acts as a preattentive cue that facilitates segregation of harmonically unrelated frequency components (Alain *et al.*, 2001). If the segregation of a mistuned harmonic involves low-level processing it might be expected to be present early in development. No research to date has used the mistuning paradigm to address infants' sensitivity to harmonic structure. Here we use a conditioned head-turn procedure (Werker *et al.*, 1997) to measure infants' ability to detect a mistuning in the third harmonic of a six-component tone in comparison to that of adults. The conditioned head-turn procedure is ideal for testing auditory perception in infants around 6 months of age because they possess adequate control over head movements and are entertained by the reinforcers (Werker *et al.*, 1997).

II. METHOD

A. Participants

Ten adult listeners (4 males, mean age = 24.4 years ± 1.88) with no reported hearing impairments and twenty-four 6-month-old infants (11 males, mean age = 6.28 months ± 0.20) participated. Five infants were excluded because they did not pass initial training and 1 because of equipment failure. An additional 21 infants were excluded because they were too tired and fussy to complete the second experimental block. All infants were screened at birth to ensure no sensorineural hearing loss according to the Ontario Infant Hearing Program (Hyde, 2005). After giving informed consent to participate, parents of the infants were asked to fill out a brief questionnaire for auditory screening purposes. According to the questionnaire, no infants had a history of frequent ear infections, pressure-equalizing tubes, or hearing impairment in the family and all were healthy at the time of testing.

B. Stimuli

Stimuli were created using ADOBE AUDITION 6.0. The in-tune complex tone was 500 ms in duration including 50 ms rise and fall times, had a pitch of 240 Hz, and contained the first six harmonics (240, 480, 720, 960, 1200, and 1440 Hz) in random phase with a 6 dB/octave roll off. The mistuned sounds were created by mistuning the third harmonic (720 Hz) of the in-tune complex upward by 1%, 2%, 4%, 6%, or 8% (e.g., in the 8% mistuned condition, the third harmonic was equal to 720 Hz × 1.08 = 777.6 Hz). Because previous literature has not addressed whether young infants are able to discriminate any amount of mistuning of any harmonic, we chose stimulus parameters that gave rise to a clear, salient perception of two auditory objects in adults (Hartman *et al.*, 1990; Lin and Hartman, 1998; Moore *et al.*, 1986).

C. Infant procedure

Parents sat in the sound attenuated chamber (Industrial Acoustics Company, Bronx, NY) with their infant sitting on their lap facing the experimenter. Sounds were presented by a Macintosh (Cupertino, CA) G4 computer through an NAD C352 amplifier to a custom built audiological speaker (Audio Visual Methods, Toronto, Canada) located to the left of the infant. The experimenter sat behind a small desk that housed several stuffed toys and a button box out of the infant's view. The button box was connected to the computer through a custom-built interface to a National Instruments (Austin, TX) PCI-DIO96 I/O card. Under the speaker to the left of the infant was a cabinet with four compartments that each housed a mechanical toy and lights used to reward the infant for correct responses. These toys and lights were also controlled by the computer through the custom-built interface. The cabinet had a smoked Plexiglas™ front such that each mechanical toy was not visible unless the lights in that compartment were illuminated. The parent and experimenter listened to continuous music through headphones that masked the stimuli the infant was hearing in order to ensure that they could not bias the infant's behavior.

The in-tune stimulus repeated continuously with a stimulus onset asynchrony of 1500 ms at approximately 70 dB(A), measured at the location of the infant's head, over a noise floor of 26 dB(A). When the infant was attentive and facing forward (the experimenter attracted the infant's attention with stuffed toys if necessary), the experimenter pressed a button that signaled to the computer to present a trial. On half of the trials, the mistuned complex replaced one instance of the in-tune complex (change trials) and on the other half, the in-tune complex was presented (control trials). The experimenter was not aware of whether a change or control trial was presented or exactly when a trial was presented. The experimenter pressed a second button on the button box whenever the infant turned his or her head at least 45° toward the speaker following the presentation of a trial. The computer kept track of head turns. If the infant turned his or her head within a 2000 ms window post onset of a change trial, the computer rewarded the infant by turning on the lights and a mechanical toy in one of the four compartments of the toy cabinet. Each experimental phase (condition) consisted of 24 trials, 12 control and 12 change trials, presented in quasi-random order such that no more than 3 control trials were presented in a row. Each experimental phase was preceded by a training phase in which the intensity of the mistuned complexes was 6 dB higher [76 dB(A)] than the in-tune complexes in order to teach infants that head turns to a change in stimulus were rewarded with animated toys. During training there were no control trials, and thus only hits and misses were recorded. In order to proceed to the experimental phase, infants were required to successfully turn toward the speaker on four consecutive change trials within 20 trials. If this criterion was not met, testing ended, and the infant's data were not used. During testing, all stimuli were presented at 70 dB(A).

Because young infants can only remain attentive and cooperative for a short time, we were only able to test each infant in two mistuning conditions. All infants were run on the 8% mistuning condition first. If the infant completed the 8% condition and was not fussy, then they began the training for *one* of the 6%, 4%, or 2% conditions. Eight infants completed testing in each of the groups (A: 8% and 6%; B: 8% and 4%; and C: 8% and 2%). For each condition, the data of interest were the number of hits (correct turns to a mistuned complex) and false alarms (incorrect turns to an in-tune complex) during the testing phase.

D. Adult procedure

Adults were tested individually in as similar a manner to infants as possible. Adults were seated in the sound attenuated chamber directly across from the experimenter. The stimulus presentation and procedure were the same as with infants, except that adults raised a hand instead of turning to indicate when they heard two objects. Unlike infants, however, each adult was tested in all mistuning conditions (8%, 6%, 4%, 2%, and 1%).

III. RESULTS

A. Infants

A d' score was calculated for each infant for each condition. Individual d' values were calculated to avoid the underestimation of sensitivity that can result from calculating d' based on average hit and false alarm rates from individuals with similar d' scores, but different response biases (Macmillan and Creelman, 2005). A correction was applied to adjust for perfect proportions in which proportions of 1 and 0 (representing infinite d' values) were converted to $1 - 1/(2N)$ and $1/(2N)$, respectively, where N is the number of trials on which the proportion is based (Macmillan and Creelman, 2005). All three groups of infants performed similarly in the initial 8% mistuning condition and a one-way analysis of variance (ANOVA) showed no significant differences across the groups, $F(2,21) = 0.24$, $p = 0.79$. A second one-way ANOVA comparing performance in the 6%, 4%, and 2% mistuning conditions showed a significant effect of condition, $F(2, 21) = 17.34$, $p < 0.001$. *Post hoc* tests indicated that performance in the 2% mistuning condition was significantly worse than in both the 4% ($p < 0.001$) and 6% ($p < 0.001$) conditions, which did not differ significantly ($p = 0.64$). T-tests on each condition revealed that performance was significantly above chance levels at 8% ($p < 0.001$), 6% ($p = 0.001$), and 4% ($p < 0.001$) mistunings, but not at 2% ($p = 0.60$; Fig. 1).

B. Adults

Using d' scores as the dependent measure, adults performed significantly above chance at all mistuning levels (all

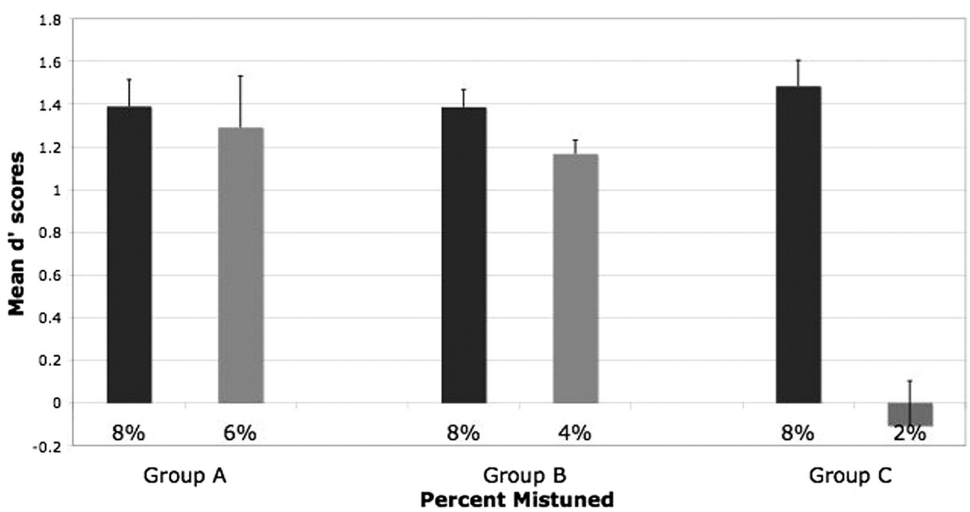

FIG. 1. Infant discrimination of the mistuned third harmonic. Group A performed equally well at 8% and 6% mistunings. Group B preformed equally well at 8% and 4% mistunings. Group C performed significantly better at 8% than 2% mistunings. Infants were significantly above chance levels at discriminating the 8%, 6%, and 4% mistunings but not the 2% mistuning. Error bars reflect standard error of the mean.

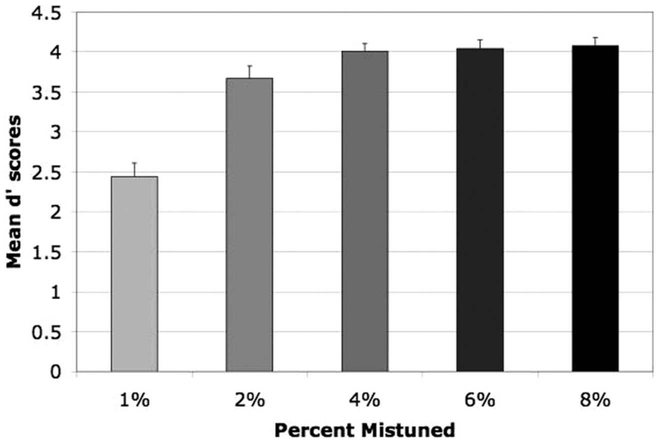

FIG. 2. Adult discrimination of the mistuned third harmonic. Adults performed well at all levels of mistuning, although significantly worse at 1% compared to all other levels of mistuning tested. Error bars represent standard error of the mean.

$ps < 0.001$). A one-way ANOVA revealed a significant effect of mistuning condition, $F(4,\ 22) = 10.421$, $p < 0.05$. Tukey's *post hoc* tests revealed that performance in the 1% mistuning condition was significantly worse compared to all other conditions (Fig. 2).

IV. DISCUSSION

The results indicate that infants are sensitive to violations in harmonic structure at 6 months of age. Specifically, infants are able to discriminate when the third harmonic in a six-harmonic complex tone is mistuned, with a threshold between 2% and 4%. This threshold is in line with infant discrimination abilities as measured for pure tones (Werner, 2007) and guitar tones (Trainor *et al.*, 2011). Adults tested with a similar procedure showed thresholds below 1% (also in line with reported pitch discrimination abilities; Olsho, 1984), suggesting that sensitivity to harmonic structure improves with age. Previous work indicates that 4-month-old infants can integrate harmonics in order to perceive the pitch of a complex tone even when the fundamental is missing (He *et al.*, 2009). The current study suggests that infants are also able to successfully segregate components that do not fit the harmonic structure of a complex sound, provided that the deviation is large enough. This suggests that infants can use harmonicity cues to detect multiple simultaneous auditory sound sources or auditory objects, although further experiments are needed to verify whether infants actually hear separate objects.

The ability to process the harmonic structure of complex sounds is important for reasons that extend beyond the musical domain. Infants' ability to segregate the auditory scene into auditory objects (or streams) corresponding to sound sources in their auditory worlds is critical for the remarkable language learning that occurs within the first years of life. Although several studies have demonstrated infants' ability to perceptually organize sequential, non-overlapping stimuli (Demany, 1982; McAdams and Bertoncini, 1997; Smith and Trainor, 2011), virtually all real-world auditory environments consist of simultaneous, overlapping sounds (e.g., multiple talkers, or a single talker amid extraneous environmental sounds). For this reason, the present examination of the perception of simultaneous auditory objects provides a point of departure for increased understanding of how infants make sense of the noisy world into which they are born.

Abdala, C., and Chatterjee, M. (**2003**). "Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans," J. Acoust. Soc. Am. **114**, 932–943.

Abdala, C., and Folsom, R. C. (**1995**). "The development of frequency resolution in humans as revealed by the auditory brian-stem response recorded with notched-noise masking," J. Acoust. Soc. Am. **98**, 921–930.

Alain, C., Arnott, S. T., and Picton, T. W. (**2001**). "Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials," J. Exp. Psychol. Hum. Percept. Perform. **27**, 1072–1089.

Alain, C., and McDonald, K. L. (**2007**). "Age-related differences in neuromagnetic brain activity underlying soncurrent sound perception," J. Neurosci. **27**, 1308–1314.

Alain, C., McDonald, K. L., Ostroff, J. M., and Schneider, B. (**2001**). "Age-related changes in detecting a mistuned harmonic," J. Acoust. Soc. Am. **109**, 2211–2216.

Alain, C., and Schuler, B. M. (**2002**). "Neural activity associated with distinguishing concurrent auditory objects," J. Acoust. Soc. Am. **111**, 990–995.

Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., and Taylor, M. (**2003**). "Developmental changes in distinguishing concurrent auditory objects," Brain Res. Cognit. Brain Res. **16**, 210–218.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sounds* (MIT Press, Cambridge, MA), pp. 1–45.

Demany, L. (**1982**). "Auditory stream segregation in infancy," Infant Behav. Dev. **5**, 261–276.

Folsom, R. (**1985**). "Auditory brainstem responses from human infants: Pure-tone masking profiles for clicks and filtered clicks," J. Acoust. Soc. Am. **78**, 555–562.

Folsom, R., and Wynne, M. (**1987**). "Auditory brainstem responses from human adults and infants: wave-V tuning curves," J. Acoust. Soc. Am. **81**, 412–417.

Hartmann, W. M. (**1996**). "Pitch periodicity, and auditory organization," J. Acoust. Soc. Am. **100**, 3491–3502.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). "Hearing a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712–1724.

He, C., Hotson, L., and Trainor, L. J. (**2007**). "Mismatch responses to pitch changes in early infancy," J. Cogn. Neurosci. **19**, 878–892.

He, C., Hotson, L., and Trainor, L. J. (**2009**). "Maturation of cortical mismatch responses to occasional pitch change in infancy: Effects of presentation rate and magnitude of change," Neuropsychologia **47**, 218–229.

Hyde, M. L. (**2005**). "Newborn hearing screening programs: Overview," J. Otolaryngol. **34**, S70–S78.

Lin, J.-Y., and Hartmann, W. M. (**1998**). "The pitch of a mistuned harmonic: Evidence for a template model," J. Acoust. Soc. Am. **103**, 2608–2617.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A Users Guide*, 2nd ed. (Erlbaum, Mahwah, NJ), pp. 8–9 and 332.

Maxon, A. B., and Hochberg, I. (**1982**). "Development of psychoacoustic behavior: Sensitivity and discrimination," Ear. Hear. **3**, 301–308.

McAdams, S., and Bertoncini, J. (**1997**). "Organization and discrimination of repeating sound sequences by newborn infants," J. Acoust. Soc. Am. **102**, 2945–2953.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2010**). "Individual differences reveal the basis of consonance," Curr. Biol. **20**, 1035–1041.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

Olsho, L. W. (**1984**). "Infant auditory discrimination," Infant Behav. Dev. **7**, 27–35.

Olsho, L. W. (**1985**). "Infant auditory perception: Tonal masking," Infant Behav. Dev. **8**, 371–384.

Olsho, L. W., Koch, E. G., and Halpin, C. F. (**1987**). "Level and age effects in infant frequency discrimination," J. Acoust. Soc. Am. **82**, 454–464.

Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., and Sperduto, V. (**1982**). "Auditory frequency discrimination in infancy," Dev. Psychol. **18**, 721–726.

Plomp, R., and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**, 548–560.

Smith, N. A., and Trainor, L. J. (**2011**). "Auditory stream segregation improves infants' selective attention to target tones amid distractors," Infancy, doi: 10.1111/j.1532-7078.2011.00067.x.

Spetner, N. B., and Olsho, L. W. (**1990**). "Auditory frequency resolution in human infancy," Child Dev. **61**, 632–652.

Swoboda, P. J., Morse, P. A., and Leavitt, L. A. (**1976**). "Continuous vowel discrimination in normal and at-risk infants," Child Dev. **47**, 459–465.

Teas, D. C., Klein, A. J., and Kramer, S. J. (**1982**). "An analysis of auditory brainstem responses in infants," Hear. Res. **7**, 19–54.

Trainor, L. J., and Corrigall, K. A. (**2010**). "Music acquisition and effects of musical training," in *Springer Handbook on Music Perception*, edited by M. Riess-Jones (Springer, New York), pp. 89–127.

Trainor, L. J., and He, C. (**2011**). "Auditory and musical development," in *Oxford Handbook of Developmental Psychology*, edited by P. Zelazo (Oxford University Press, Oxford, UK).

Trainor, L. J., Lee, K. L., and Bosnyak, D. J. (**2011**). "Cortical plasticity in 4-month-old infants: Specific effects of experience with musical timbres," Brain Topogr. **24**, 192–203.

Werker, J. F., Polka, L., and Pegg, J. E. (**1997**). "The conditioned head turn procedure as a method for testing infant speech perception," Early Dev. Parent. **6**, 171–178.

Werner, L. A. (**2007**). "Human Auditory Development," in *The Senses: A comprehensive Reference*, Vol. 3, edited by R. Hoy, P. Dallos, and D. Oertel (Academic, New York), pp. 871–874.