

## RESPONSE

# Why voice melody alone cannot explain neonates' preference for speech

Athena Vouloumanos<sup>1</sup> and Janet F. Werker<sup>2</sup>

1. Department of Psychology, McGill University, Canada

2. Department of Psychology, University of British Columbia, Canada

This is a response to the commentary on Vouloumanos and Werker (2007) by Rosen and Iverson (2007).

Are humans born with a bias for listening to the vocalizations of their species? In Vouloumanos and Werker (2007, this issue), we present data demonstrating that from birth, the human infant prefers listening to speech, compared with non-speech sounds that mimic spectral and temporal properties of speech. Rosen and Iverson (2007, this issue) criticize this interpretation, first arguing that the preference we have shown is based on voice melody rather than speech *per se*; second, they argue that such a voice melody preference likely stems from prenatal learning, rather than from an innate bias – a claim we didn't make in Vouloumanos and Werker, but that is addressed by new data we present here.

Turning to Rosen and Iverson's first point, although we agree that a voice melody account of newborns' preference for speech is not altogether impossible, we find it implausible. Voice melody, or pitch, is the subjective highness or lowness of a sound as perceived by the human ear. Although pitch extraction is not fully understood (e.g. Patel & Balaban, 2001), in natural speech, pitch is generally perceived as the fundamental frequency (F0) of an utterance, which is the frequency with which a particular speaker's vocal folds vibrate (typically around 200 Hz for a woman's voice). Because of resonance properties of sound, F0 is reflected in 'harmonics' at integer multiples of F0 (e.g. an F0 of 150Hz (itself the first harmonic), will have harmonics at 300 Hz, 450 Hz, 600 Hz, etc.), which contribute to the perception of pitch if, for example, F0 is missing. Research on infant pitch perception is limited, but has shown that 7-month-old infants demonstrate some adult-like characteristics in their perception of pitch (Montgomery & Clarkson, 1997). Even at this age, however, there is considerable variation in individual infants' abilities to recover pitch when F0 is missing (Clarkson, 1992). Pitch extraction in younger

infants is currently poorly understood but is believed to differ from adult pitch perception (Bundy, Colombo & Singer, 1982; Clarkson, 1992). Though neonates are sensitive to pitch contours, discriminating, for example, high-low pitch from low-high pitch in bimoraic stimuli (Nazzi, Floccia & Bertoncini, 1998), the mechanism of pitch extraction in neonates has not been investigated.

To examine neonates' preference for speech, the non-speech sounds we used were a variant on sine-wave analogues (SWA) of speech (Remez, Rubín, Pisoni & Carrell, 1981). SWA consist of time-varying sinusoidal waves, or sine waves, that track the centre frequencies of the energy bands (formants) of natural speech to reproduce the changes in these frequency peaks across time. SWA are typically composed of three sinusoidal waves that reproduce the changes in the first three formants of speech so adroitly that under the right circumstances, adult listeners perceive SWA as intelligible (if weird) speech (Remez *et al.*, 1981). At stake here is which component of our SWA conveyed the perception of pitch. Rosen and Iverson suggest that because the first formant (F1) is usually heard as conveying pitch in SWA, even with the addition of F0 (Remez & Rubín, 1984), F1 is likely to convey perceived pitch in our stimuli as well, and thus, the voice melody perceived in our SWA is less salient compared to that in the speech set. We would argue that the F0 component in our SWA was salient, and that it, rather than F1, accounted for the perceived pitch. The key lies in the construction of the stimuli by Sonya Bird and Guy Carden, of the University of Victoria, and the University of British Columbia, respectively. While creating the SWA, they found that the first three formants (F1, F2, and F3) were virtually identical across the multiple natural speech tokens. For this reason, they selected *one* representative set of the first three formants

Address for correspondence: Athena Vouloumanos, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada; e-mail: athena.vouloumanos@mcgill.ca

from *one* token, and created the F1-F2-F3 sine-wave complex from this one token. The four different SWA used in the study were then created by superimposing a sine wave tracking the F0 of the four natural infant-directed speech tokens onto this *single* F1-F2-F3 complex. Inasmuch as the listener can hear any difference between the different SWA tokens, this difference is specified *entirely* by F0, because it is the only component that differs between the tokens. Even the most casual listener presented with the different non-speech tokens nonetheless hears them as distinct (non-speech tokens can be heard at <http://www.phon.ucl.ac.uk/reports/DevScience2006/>). The pitch contour that conveys voice melody in the sine-wave analogues can be heard readily, requiring neither careful nor analytic listening. A preference for speech is thus unlikely to be captured entirely by a preference for the voice melody of speech.

Second, independent of whether the non-speech stimuli capture voice melody to the same extent that natural speech tokens do, the prenatal environment is unlikely to provide the kind of information that Rosen and Iverson claim it does with respect to voice melody in these two sets of sounds. While we made no claim about innateness in Vouloumanos and Werker (2007), we more carefully address the kind of information available prenatally in a follow-up (unreported) experiment. In this experiment, we low pass filtered (LPF) the speech and SWA sounds with a 400-Hz filter to emulate what a foetus would be likely to hear (Abrams & Gerhardt, 2000). This frequency range is sufficient to convey information about the mother's voice (Spence & Freeman, 1996) and about the native language (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison, 1988), and is consistent with human post-natal preferences (DeCasper & Fifer, 1980; Moon, Cooper & Fifer, 1993). We tested whether neonates could discriminate between LPF speech and LPF non-speech using the Cowan method (Cowan, Suomi & Morse, 1982), which compares infants' habituation slopes for different types of stimuli (Floccia, Nazzi & Bertoncini, 2000). When discriminable stimuli are presented in alternating minutes, newborns will maintain their high amplitude sucking rate, whereas when stimuli are non-discriminable, they are treated as a single repeating stimulus, and newborns' sucking rates decrease significantly. If newborns treat LPF speech and LPF SWA as discriminable stimuli, they should maintain their sucking rate. If, however, LPF speech and LPF SWA are not discriminable, newborns should show a significant decrease in their sucking rate. Pilot data are clear: When we present these two alternating sets of LPF sounds to newborns, their high amplitude sucking rate decreases significantly, suggesting that neonates cannot discriminate between LPF speech and LPF SWA. This suggests that the

information required to discriminate between SWA and bona fide speech is contained in higher frequencies which are severely attenuated, if available at all, in the prenatal environment. In short, whatever aspect of voice melody infants are familiar with prenatally is not likely to be sufficient to discriminate between our speech and SWA tokens, and thus prenatal familiarity with voice melody *per se* is unlikely to account for neonates' preference for speech.

In addition to confirming that voice melody available prenatally is indistinguishable between speech and our sine-wave stimuli, and thus is unlikely to account for post-natal preferences, these new data reduce the range of plausible roles for human prenatal listening experience in the preference for speech over sine-wave analogues reported in Vouloumanos and Werker (2007). This suggests the intriguing possibility that human neonates' preference for speech could be innate.

## Acknowledgements

We thank Stuart Rosen and Evan Balaban for useful discussions, and Gary Marcus for comments on an earlier draft. Research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant 312281-05 (AV), and NSERC Discovery grant RGP81103, the Human Frontiers Science Program, and a Canada Research Chair (JFW).

## References

- Abrams, R.M., & Gerhardt, K.J. (2000). The acoustic environment and physiological responses of the fetus. *Journal of Perinatology*, **20** (8 Pt 2), S31–S36.
- Bundy, R.S., Colombo, J., & Singer, J. (1982). Pitch perception in young infants. *Developmental Psychology*, **18** (1), 10–14.
- Clarkson, M.G. (1992). Infants' perception of low pitch. In L.A. Werner & E.W. Rubel (Eds.), *Developmental psychoacoustics* (pp. 159–188). Washington, DC: American Psychological Association.
- Cowan, N., Suomi, K., & Morse, P.A. (1982). Echoic storage in infant perception. *Child Development*, **53** (4), 984–990.
- DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, **208** (4448), 1174–1176.
- Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, **3** (3), 333–343.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, **29** (2), 143–178.
- Montgomery, C.R., & Clarkson, M.G. (1997). Infants' pitch perception: masking by low- and high-frequency noises.

- Journal of the Acoustical Society of America*, **102** (6), 3665–3672.
- Moon, C., Cooper, R.P., & Fifer, W.P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, **16** (4), 495–500.
- Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development*, **21** (4), 779–784.
- Patel, A.D., & Balaban, E. (2001). Human pitch perception is reflected in the timing of stimulus-related cortical activity. *Nature Neuroscience*, **4** (8), 839–844.
- Remez, R.E., & Rubin, P.E. (1984). On the perception of intonation from sinusoidal sentences. *Perception and Psychophysics*, **35** (5), 429–440.
- Remez, R.E., Rubin, P.E., Pisoni, D.B., & Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science*, **212** (4497), 947–949.
- Rosen, S., & Iverson, P. (2007). Constructing adequate non-speech analogues: what *is* special about speech anyway? *Developmental Science*, **10** (2), 165–169.
- Spence, M.J., & Freeman, M.S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development*, **19** (2), 199–212.

Received 19 October 2005

Accepted: 1 February 2006