

# Audible Television and Decreased Adult Words, Infant Vocalizations, and Conversational Turns

## A Population-Based Study

Dimitri A. Christakis, MD, MPH; Jill Gilkerson, PhD; Jeffrey A. Richards, MA; Frederick J. Zimmerman, PhD; Michelle M. Garrison, PhD; Dongxin Xu, PhD; Sharmistha Gray, PhD; Umit Yapanel, PhD

**Objective:** To test the hypothesis that audible television is associated with decreased parent and child interactions.

**Design:** Prospective, population-based observational study.

**Setting:** Community.

**Participants:** Three hundred twenty-nine 2- to 48-month-old children.

**Main Exposures:** Audible television. Children wore a digital recorder on random days for up to 24 months. A software program incorporating automatic speech-identification technology processed the recorded file to analyze the sounds the children were exposed to and the sounds they made. Conditional linear regression was used to determine the association between audible television and the outcomes of interest.

**Outcome Measures:** Adult word counts, child vocalizations, and child conversational turns.

**Results:** Each hour of audible television was associated with significant reductions in age-adjusted *z* scores for child vocalizations (linear regression coefficient,  $-0.26$ ; 95% confidence interval [CI],  $-0.29$  to  $-0.22$ ), vocalization duration (linear regression coefficient,  $-0.24$ ; 95% CI,  $-0.27$  to  $-0.20$ ), and conversational turns (linear regression coefficient,  $-0.22$ ; 95% CI,  $-0.25$  to  $-0.19$ ). There were also significant reductions in adult female (linear regression coefficient,  $-636$ ; 95% CI,  $-812$  to  $-460$ ) and adult male (linear regression coefficient,  $-134$ ; 95% CI,  $-263$  to  $-5$ ) word count.

**Conclusions:** Audible television is associated with decreased exposure to discernible human adult speech and decreased child vocalizations. These results may explain the association between infant television exposure and delayed language development.

*Arch Pediatr Adolesc Med.* 2009;163(6):554-558

**T**ELEVISION VIEWING DURING very early childhood is a growing but understudied phenomenon.<sup>1-6</sup> The American Academy of Pediatrics discourages television or video viewing before the age of 2 years, suggesting instead that parents focus on interactive play to foster appropriate child development.<sup>6</sup>

Language acquisition is a critical developmental task in early childhood that is promoted by certain activities, including interacting with adults.<sup>7-9</sup> In a prior study, we found an association between infant television or video viewing and delayed language development.<sup>10</sup> What factors might mediate this association is not entirely clear, however. One small laboratory-based study found that parents interact less with their children in the presence of a television set that is turned on.<sup>11</sup> In a separate retrospective study conducted in a low-income population, tele-

vision exposure time was associated with self-reported decreased parent-child vocal interactions.<sup>12</sup> To date, no study has prospectively examined the effects of child television viewing on the frequency and nature of adult-child interactions in a population-based sample outside of a laboratory setting. We hypothesized that television exposure would be associated with decreased adult and child vocal activity.

## METHODS

### DATA SOURCE

Data for this study were obtained from the LENA Foundation Natural Language Study.<sup>13</sup> LENA is a language environment analysis system (LENA Foundation, Boulder, Colorado) designed to provide parents, clinicians, and researchers with information about the language environment of infants and toddlers. The LENA system contains a digital language pro-

**Author Affiliations:** Center for Child Health, Behavior and Development; and Department of Health Services, University of Washington (Dr Christakis); Seattle Children's Research Institute (Drs Christakis, Zimmerman, and Garrison), Seattle, Washington; and LENA Foundation, Boulder, Colorado (Drs Gilkerson, Xu, Gray, and Yapanel, and Mr Richards).

**Table 1. Fidelity Matrix of Human Transcription and Machine Coding<sup>a</sup>**

	% Machine-Coded as . . . <sup>b</sup>			
	Adult Speech	Child Vocalization	Television	Other
Among segments human-transcribed as . . .				
Adult speech	82.0	1.9	3.9	12.2
Child vocalization	7.3	76.0	0.1	16.6
Television	7.8	0.5	70.5	21.2
Other	13.5	4.5	6.3	75.7
	% Among Segments Machine-Coded as . . . <sup>c</sup>			
	Adult Speech	Child Vocalization	Television	Other
Human-coded as . . .				
Adult speech	67.9	4.1	13.0	5.7
Child vocalization	2.5	69.8	0.2	3.2
Television	0.9	0.2	33.4	1.4
Other	28.7	25.9	53.4	89.7

<sup>a</sup>Human transcription was considered the criterion standard.

<sup>b</sup>Sensitivity.

<sup>c</sup>Specificity.

cessor that children (aged 2-48 months) wear in the pocket of clothing custom made for the device. It records everything the child says and hears during a continuous 12- to 16-hour day. The audio data are transferred to a computer and analyzed by the LENA language environment analysis software.

The LENA software contains advanced speech-identification algorithms that automatically generate language environment information based on models of the following segments: adult male, adult female, key child, other child, overlapping speech, noise, silence, and television/electronic sound. For a detailed overview of the LENA system software, please see LENA Foundation technical reports.<sup>14</sup>

The system has been tested and validated regarding the software's estimates of adult words, child vocalizations, conversational turns, and the presence of audible electronic media. Compared with trained human transcribers, the software achieves a high degree of fidelity in coding. Among segments that human transcribers identified as adult speech, 82% were correctly identified as such by the software, with less than 2% erroneously coded as child speech, 4% coded erroneously as television, and 12% erroneously coded as other (overlap, vegetative sounds, cries, distant television, etc) (**Table 1**). Among segments that the software identified as adult speech, 68% were also so identified by human transcribers, with most of the rest (29% of the total) identified by human transcribers as other sound, mostly as overlap between adult speech and sound from another source. Overlap occurs when the system (or the human transcriber) cannot identify a dominant sound; as such, from the perspective of the child, it might rightly be appreciated as noise. The full fidelity matrix is reported in Table 1. The overall fidelity is very good, with diagonal elements generally exceeding 70%, indicating a high degree of concordance between machine-coded identification and identification by human transcribers. More importantly, when miscoding did occur, it rarely involved confusion of key variables in this analysis, for example, adult speech and child vocalizations or adult speech and television.

The fidelity matrix assumes that human assignment of a sound's origin is the gold standard, an assumption that may or may not be valid when tested against direct, live observation. We therefore also calculated Cohen  $\kappa$  statistics for the identification of adult and child speech and television between human coders and the LENA software. The  $\kappa$  values range between 0.43 and 0.70, showing moderate to good agreement.

## STUDY PARTICIPANTS

Between January and June of 2006, parents of infants and toddlers (aged 2-48 months) were recruited through advertisements in local newspapers and direct mail solicitation. Baseline demographic data were used to select a representative sample. The recruitment goal was to achieve a final study sample of at least 300 children that matched US census data for maternal education and child sex, both for the overall sample and within each single-month age stratum from age 2 months to 36 months (for a total of 35 distinct age strata). As study enrollment progressed, some children (n=15) aged to 37-48 months at the time of their first recording sample. Children were excluded if they had a language delay diagnosed or if the primary language spoken at home was not English. Informed consent was obtained from the parents. The study protocol was reviewed and approved by Essex Institutional Review Board (Lebanon, New Jersey).

## DATA COLLECTION

At enrollment, parents completed background questionnaires, which included extensive demographic questions. To simplify data collection, parents were assigned a random day of the month ranging from 1 to 29 and asked to record on that day each month. Just prior to their recording day, parents were express-mailed the LENA digital language processor, a small device that records every sound within earshot of the child for a 12- to 16-hour period. These recording sessions occurred monthly; the total number of months of participation varied from 1 to 24, with a median study participation of 6 months.

The digital language processor weighs about 2 oz and is worn in a front pocket of a specially designed vest for young children. Parents were instructed to begin recording the moment the child awoke and until he or she went to bed at night, removing the device only during naps, baths, and car rides. Parents were instructed to behave as they normally would, with the exception that during the first 3 months they should turn off any ambient noise (eg, television and radio); for the final 3 months, they were told that there was no need to do this. This request was made because it was unclear whether ambient noise

would interfere with recording sessions. Once initial data were analyzed, it became apparent they would not.

## OUTCOME VARIABLES

Adult word count estimates were calculated based on analysis of the adult male and adult female segments. The adult word count report uses acoustic features to estimate the number of adult words spoken near the child with high accuracy compared with human transcribers.<sup>15</sup> Child vocalization estimates were calculated based on the number of vocalizations the child made during the session. A child vocalization is defined as a segment of meaningful child speech (ie, excluding child nonspeech: fixed signals [eg, cries] and vegetative sounds) of any length surrounded by 300 milliseconds or more of nonspeech or silence. Child vocalization outcomes included counts (total number of vocalizations during the recording session) and duration (total time of vocalizations during the session). In validating the LENA software, 70 different children aged 2 to 36 months were tested with 1 hour of audio data. Conversational turns were calculated as the total number of times per session the child engaged in vocal interaction with an adult. It is an estimate of the number of back-and-forth interactions between the key child wearing the LENA device and an adult. For example, 1 conversational turn is counted if the child vocalizes and an adult responds (or vice versa) within 5 seconds. Because child vocalizations and conversational turns vary significantly by age, the outcome used in the regression model was the age-specific *z* score for child vocalizations and conversational turns.

## TELEVISION

The LENA software automatically identifies electronic media segments based on acoustic modeling of sounds transcribed by humans as television or radio; for simplicity, we refer to these segments as simply *television*. The identification of television segments occurs in 2 steps. First, all recorded sound is compared with a general television model, and those segments that match this model are assigned to the general television category. In a more strict way, under a maximum likelihood framework, all recorded sound is decoded into the segments of different sound categories based on the models of the categories. Second, all segments identified as television are compared with the silence model to generate a likelihood ratio test. Segments that resemble the silence model are considered faint/unclear and the rest are labeled clear television segments. Adult speech and child vocalizations are processed with the same likelihood ratio test procedure to identify clear segments from faint ones.

## STATISTICAL ANALYSIS

We conducted a fixed-effect analysis using linear regression models with child as a fixed effect, so that each child served as his or her own control. That is, we exploited the natural variation with each child's daily television viewing to understand its relationship to adult word count and conversational turn estimates. This method explicitly controls for all of the characteristics at the child and family level that may confound the relationships between our primary predictor variable and outcome of interest. Essentially, it compares the amount of vocalizations and conversational turns that an individual child experiences on high-television exposure days with low-television exposure days. Accordingly, the only variables we included as covariates were those that could be expected to vary within a child over time—in this case, television exposure, the time of day that the recording session began, the recording session number for the child (first, second, etc), and the length of each recording session.

## RESULTS

One thousand nine hundred ninety-eight potential participants responded to the recruitment advertisement. From these, 435 potential participants were selected based on demographic variables. Parents of 364 of these children reviewed the consent form and 334 enrolled. There were 329 participants who contributed at least 1 recording with usable speech data. The mean number of recorded sessions per child was 8.2 with a range of 1 to 24. Demographic data on the included sample are presented in **Table 2**.

In the regression analyses, television exposure was associated with significantly reduced child vocalization and adult word counts (**Table 3**). Every additional hour of television exposure during a recording session was associated with a decrease of 0.26 in the *z* score for child vocalizations; similar results were observed for child vocalization duration (−0.24) and conversational turns (−0.22). As is shown in **Figure 1**, the slope associated with these effects was very similar across the 3 child vocalization outcomes.

Likewise, each additional hour of television exposure was associated with a decrease of 770 in the number of words the child heard from an adult during the recording session, which represents a 7% decrease. Just as a greater proportion of a child's total exposure to adult speech was from women, so too was a greater decrease observed in female word counts with each hour of television exposure; this can be seen in **Figure 2**, where the slope associated with female word count is similar to that for total adult word count.

## COMMENT

We found that having a television on was associated with significant reductions in discernible parental word counts, child vocalizations, and conversational turns for children 2 to 48 months of age. Some of these reductions are likely due to children being left in front of the television screen, but others likely reflect situations in which adults, though present, are distracted by the screen and not interacting with their infant in a discernible manner. At first blush, these findings may seem entirely intuitive. That is, parents engage their infants less when the television is on. However, these findings must be interpreted in light of the fact that purveyors of infant DVDs claim that their products are designed to give parents and children a chance to interact with one another, an assertion that lacks empirical evidence.<sup>16</sup> Furthermore, given that 30% of households have televisions on all of the time, our results beg the question of how many opportunities of child and parent vocalizations are being displaced.<sup>5</sup>

The magnitude of our findings is significant as well. The effect size, per hour of television, is about one-fourth of a standard deviation for vocalization number, duration, and conversational turns. In terms of adult word count, we found that there were 500 to 1000 fewer adult words spoken per hour of television. Normative data for adult word counts indicate that adults utter approximately 941 words per hour,<sup>17</sup> suggesting that their talking is also significantly reduced when a television is audible to the child.

**Table 2. Characteristics of 329 Study Participants**

Characteristic	Percentage
Male sex	51
Age at first session, mean (SD), mo	18.6 (11.3)
Age, mo	
2-12	35
13-24	31
25-48	34
No. of recording sessions, mean (SD)	8.2 (6.8)
Duration of study enrollment, mean (SD), mo	7.5 (7.1)
Race/ethnicity	
White	79
Hispanic	9
African American	4
American Indian	1
Asian/Pacific Islander	<1
Other	4
Unknown	2
Family income, \$	
<20 000	23
20 000-40 000	30
41 000-60 000	20
61 000-100 000	17
>100 000	10
Maternal education	
Some high school	14
High school or GED diploma	27
Some college or trade school	33
Undergraduate degree	18
Graduate school	8
Paternal education	
Some high school	11
High school or GED diploma	22
Some college or trade school	21
Undergraduate degree	14
Graduate school	9
Unknown	23
Age at session, mean (SD), mo	22.7 (11.1)
Age, mo	
2-12	21
13-24	36
25-36	43
Television exposure, mean (SD), h/d	1.3 (1.1)
Child vocalization, mean (SD)	
No. of vocalizations	1816 (981)
Vocalization duration, min	21.9 (12.8)
No. of conversational turns	441 (259)
Adult word counts, mean (SD)	
Male	3419 (3016)
Female	9557 (4594)
Total	12 976 (6051)

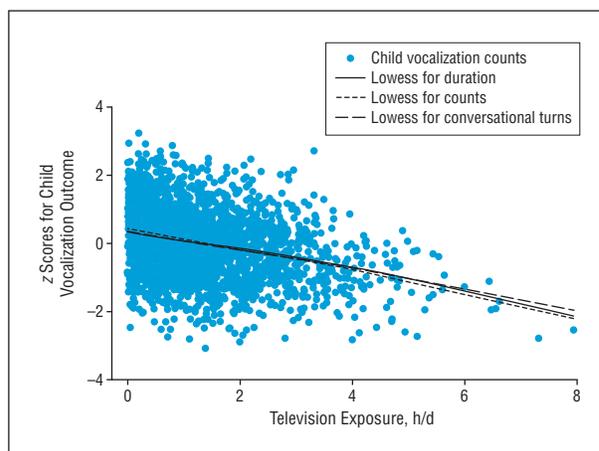
Abbreviation: GED, general educational development.

Our findings may partially explain the associations that have been previously found between infant television viewing and delayed language acquisition.<sup>10</sup> In addition, they may explain attentional and cognitive delays, as some have posited that language may be a critical mediator for both attentional capacity and thought.<sup>7</sup> Furthermore, our results highlight the need to conceptualize media exposure with consideration of more than just amount of exposure. Content has been shown to be a key mediator of effects, and these results illustrate that context (or how children watch) may also be important.<sup>18-20</sup> Given the critical role that adult caregivers play in chil-

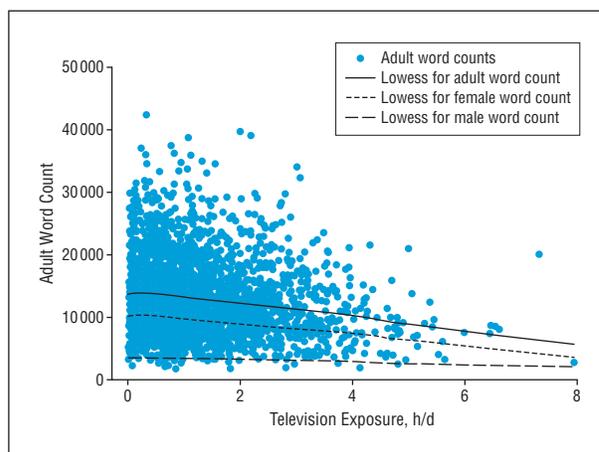
**Table 3. Television Exposure and Vocalization Conditional Linear Regression Results**

Measure	Effect (95% CI) per Hour of Television Exposure
Child vocalization age-adjusted z score	
Vocalization count	-0.26 (-0.29 to -0.22)
Vocalization duration	-0.24 (-0.27 to -0.20)
Conversational turns	-0.22 (-0.25 to -0.19)
Adult word counts	
Total	-770 (-1004 to -535)
Female adult	-636 (-812 to -460)
Male adult	-134 (-263 to -5)

Abbreviation: CI, confidence interval.



**Figure 1.** Child z scores of vocalization vs total television hours per day.



**Figure 2.** Adult word counts vs total television hours per day.

dren's linguistic development, whether they talk to their child while the screen is on may be critical and explain the effects that are attributed to content or even amount of television watched.<sup>21-23</sup> That is, whether parents talk less (or not at all) during some types of programs or at some times of the day may be as important in this age group as what is being watched.

This study has some limitations that warrant mention. First, in an observational study, causal inferences cannot be drawn. However, the statistical model we used, in which

each child functioned as his or her own control is as robust an observational design there is, because it explicitly controls for individual and familial characteristics that might bias the findings. Second, the LENA software, though extensively tested and validated, is not perfectly accurate in its determination of either television or parental or child speech. However, for this to have biased our findings, it would have to be systematically biased. There is no evidence that this is the case. Third, if the television is on and recorded by LENA, parental vocalization occurring concurrently would be coded as overlap, thereby diminishing artifactually the actual number of adult words. However, this would only happen in situations in which the parent and television are truly overlapping (and not when there is a pause in the video or television sound for parents to talk). From the perspective of an infant, this overlap still likely represents cognitive overload, because it is difficult for them to attend to 2 sounds simultaneously. Fourth, we do not know definitively if the language captured by LENA was specifically addressed to the child. Fifth, we do not know what programs were being watched and if, in some meaningful way, this mediates adult-child interactions. Finally, the data collection and the analytic plan did not distinguish between background and foreground television. That is, there is no way for us to ascertain whether or not the child was attending to the screen or that the television program was even intended for them. Accordingly, these results may misrepresent the effects of either foreground or background television, though as a summary estimate, if one had less of an effect, then the other would have to have a greater one. Furthermore, background television has, itself, been associated with decreases in children's ability to attend.<sup>24</sup>

Despite these limitations, our results have some important implications. The American Academy of Pediatrics specifically recommends against screen time for children younger than 2 years, urging more interactive play in its place.<sup>6</sup> Our results show that a trade-off between those 2 is indeed being made. Having a television on within earshot of young children diminishes their exposure to adult words, their own vocalizations, and the conversational turns in which they engage.

**Accepted for Publication:** October 30, 2008.

**Correspondence:** Dimitri A. Christakis, MD, MPH, Center for Child Health, Behavior and Development, 1100 Olive Way, Ste 500, Seattle, WA 98101 (dachris@u.washington.edu).

**Author Contributions:** Dr Christakis had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Dr Gilkerson coordinated the collection of the data, including the recruitment and management of the sample. Drs Xu, Yapanel, and Gray developed and refined the audio-processing algorithms. Mr Richards processed the data. *Study concept and design:* Christakis, Gilkerson, Zimmerman, Garrison, Gray, and Yapanel. *Acquisition of data:* Gilkerson, Richards, and Zimmerman. *Analysis and interpretation of data:* Christakis, Zimmerman, Garrison, Xu, Gray, and Yapanel. *Drafting of the manuscript:* Christakis and Garrison. *Critical revision of the manuscript for important intellectual content:* Christakis, Gilkerson, Richards, Zimmer-

man, Xu, Gray, and Yapanel. *Statistical analysis:* Christakis, Richards, Zimmerman, and Garrison. *Obtained funding:* Christakis and Zimmerman. *Administrative, technical, and material support:* Gilkerson, Xu, Gray, and Yapanel.

**Financial Disclosure:** Drs Gilkerson, Xu, Gray, and Yapanel, and Mr Richards were employed by the LENA Foundation.

**Funding/Support:** The LENA Foundation paid for the data collection.

**Additional Contributions:** We gratefully acknowledge Terrance Paul, JD, MBA, for conceiving of the LENA system and for personally funding and directing its development as well as the development of the LENA Foundation Natural Language Corpus.

## REFERENCES

1. Certain LK, Kahn RS. Prevalence, correlates, and trajectory of television viewing among infants and toddlers. *Pediatrics*. 2002;109(4):634-642.
2. Christakis DA, Zimmerman FJ. Media as a public health issue. *Arch Pediatr Adolesc Med*. 2006;160(4):445-446.
3. Christakis DA, Zimmerman FJ, Garrison MM. Television viewing in child care programs: a national survey. *Commun Rep*. 2006;19(2):111-120.
4. Zimmerman FJ, Christakis DA, Meltzoff AN. Television and DVD/video viewing in children younger than 2 years. *Arch Pediatr Adolesc Med*. 2007;161(5):473-479.
5. Vandewater EA, Rideout VJ, Wartella EA, Huang X, Lee JH, Shim MS. Digital childhood. *Pediatrics*. 2007;119(5):e1006-e1015.
6. American Academy of Pediatrics. Committee on Public Education. American Academy of Pediatrics: children, adolescents, and television. *Pediatrics*. 2001;107(2):423-426.
7. Vygotsky LS, Kozulin A. *Thought and Language*. Cambridge, MA: Massachusetts Institute of Technology; 1986.
8. Christakis DA, Zimmerman FJ, Garrison MM. Effect of block play on language acquisition and attention in toddlers. *Arch Pediatr Adolesc Med*. 2007;161(10):967-971.
9. Pinker S. *The Language Instinct*. New York, NY: Harper Collins; 1994.
10. Zimmerman FJ, Christakis DA, Meltzoff A. Associations between media viewing and language development in children under age 2 years. *J Pediatr*. 2007;151(4):364-368.
11. Tanimura M, Okuma K, Kyoshima K. Television viewing, reduced parental utterance, and delayed speech development in infants and young children. *Arch Pediatr Adolesc Med*. 2007;161(6):618-619.
12. Mendelsohn AL, Berkule SB, Tomopoulos S, et al. Infant television and video exposure associated with limited parent-child verbal interactions in low socioeconomic status households. *Arch Pediatr Adolesc Med*. 2008;162(5):411-417.
13. Gilkerson J, Richards JA. LENA Foundation Natural Language Study. Boulder, CO: LENA Foundation; 2008. [http://www.lenafoundation.org/TechReport.aspx/Natural\\_Language\\_Study/ITR-02-2](http://www.lenafoundation.org/TechReport.aspx/Natural_Language_Study/ITR-02-2). Accessed March 3, 2009.
14. Ford M, Baer CT. The LENA language environment analysis system: audio specifications of the DLP-0121. Boulder, CO: LENA Foundation; 2008. [http://www.lenafoundation.org/TechReport.aspx/Audio\\_Specifications/ITR-03-2](http://www.lenafoundation.org/TechReport.aspx/Audio_Specifications/ITR-03-2). Accessed March 3, 2009.
15. Yapanel U, Gray SS, Xu D. Reliability of the LENA language environment analysis system in young children's natural home environment. Boulder, CO: LENA Foundation; 2008. <http://www.lenafoundation.org/TechReport.aspx/Reliability/ITR-05-2>. Accessed March 3, 2009.
16. Garrison M, Christakis D. *A Teacher in the Living Room? Educational Media for Babies, Toddlers, and Preschoolers*. Menlo Park, CA: Kaiser Family Foundation; 2005.
17. Mehl MR, Vazire S, Ramirez-Esparza N, Slatcher RB, Pennebaker JW. Are women really more talkative than men? *Science*. 2007;317(5834):82.
18. Christakis DA, Zimmerman FJ. *The Elephant in the Living Room: Make Television Work for Your Kids*. Emmaus, PA: Rodale; 2006.
19. Christakis DA, Zimmerman FJ. Violent television viewing during preschool is associated with antisocial behavior during school age. *Pediatrics*. 2007;120(5):993-999.
20. Zimmerman FJ, Christakis DA. Associations between content types of early media exposure and subsequent attentional problems. *Pediatrics*. 2007;120(5):986-992.
21. Kuhl PK. Early language acquisition: cracking the speech code. *Nat Rev Neurosci*. 2004;5(11):831-843.
22. Kuhl PK, Tsao FM, Liu HM. Foreign-language experience in infancy. *Proc Natl Acad Sci U S A*. 2003;100(15):9096-9101.
23. Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992;255(5044):606-608.
24. Anderson DR, Pempek TA. Television and very young children. *Am Behav Sci*. 2005;48(5):505-522.