



Brief article

Visual speech contributes to phonetic learning in 6-month-old infants

Tuomas Teinonen^{a,b,*}, Richard N Aslin^c, Paavo Alku^d, Gergely Csibra^b^a Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, P.O. Box 9, FI-00014 Helsinki, Finland^b Centre for Brain and Cognitive Development, School of Psychology, Birkbeck, University of London, London, UK^c Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA^d Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland

ARTICLE INFO

Article history:

Received 8 October 2007

Revised 15 May 2008

Accepted 19 May 2008

Keywords:

Visual speech

Phonetic learning

Distributional information

ABSTRACT

Previous research has shown that infants match vowel sounds to facial displays of vowel articulation [Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138–1141; Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behaviour & Development*, 22, 237–247], and integrate seen and heard speech sounds [Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, 59, 347–357; Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 45, 204–220]. However, the role of visual speech in language development remains unknown. Our aim was to determine whether seen articulations enhance phoneme discrimination, thereby playing a role in phonetic category learning. We exposed 6-month-old infants to speech sounds from a restricted range of a continuum between /ba/ and /da/, following a unimodal frequency distribution. Synchronously with these speech sounds, one group of infants (the two-category group) saw a visual articulation of a canonical /ba/ or /da/, with the two alternative visual articulations, /ba/ and /da/, being presented according to whether the auditory token was on the /ba/ or /da/ side of the midpoint of the continuum. Infants in a second (one-category) group were presented with the same unimodal distribution of speech sounds, but every token for any particular infant was always paired with the same syllable, either a visual /ba/ or a visual /da/. A stimulus-alternation preference procedure following the exposure revealed that infants in the former, and not in the latter, group discriminated the /ba/–/da/ contrast. These results not only show that visual information about speech articulation enhances phoneme discrimination, but also that it may contribute to the learning of phoneme boundaries in infancy.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The function of receptive language is to recover the intended message of the speaker. Although the dominant source of information used to convey meaning in spoken language is vocal output (i.e., speech), vocal gestures also

provide visual cues from movements of the face, including the mouth, lips, and eyebrows. These visual cues are particularly useful when the auditory cues are masked by ambient noise or less than ideal articulations (e.g., during rapid speech). Thus, an ideal listener should integrate visual cues with auditory cues to optimise the reception of the speaker's intended message.

Pre-linguistic infants are confronted with considerable ambiguity about the speaker's intended message, because the way in which the sounds of a language are mapped onto the lexicon involves a complex set of language-spe-

* Corresponding author. Address: Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, P.O. Box 9, FI-00014 Helsinki, Finland. Tel.: +358 9 191 29 465; fax: +358 9 191 29 450.

E-mail address: tuomas.teinonen@helsinki.fi (T. Teinonen).

cific phonological rules that must be learned from ambient speech input. Thus, synchronous visual input may provide additional disambiguating information that complements the auditory cues. A variety of studies have shown that infants are affected by the visual cues present during speech perception. Four-month-old infants, exposed to two faces articulating vowels on a screen, look longer at the face that matches an auditory vowel played simultaneously (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999). Even 2-month-old infants can detect the correspondence between the auditorially and visually perceived speech information (Patterson & Werker, 2003). Infants are also able to associate multiple non-linguistic events received through different modalities (for a review, see Lewkowicz 2000).

Massaro (1984) showed that visual information has an effect on how adults and children categorise heard phonemes. A visual syllable presented with an ambiguous speech sound drew the perception of the auditory sound towards the visually presented syllable. Within the fuzzy logical model of perception (Massaro, 1998), this was interpreted as a general tendency for the least ambiguous source of information to have a greater influence under ambiguous circumstances. Green and Kuhl (1989) demonstrated that visual information also influences the processing of voice-onset time information and, in contrast to Massaro's explanation, interpreted their results in terms of the interactions of specific phonetic cues, irrespective of the relative weight of auditory or visual cues. In a pioneering study of auditory-visual integration, McGurk and MacDonald (1976) found that children and adults form an integrated percept of visual and auditory speech. The perceived place of articulation of the syllable /ba/ changed when a video of a person articulating /ga/ was presented in synchrony with the speech syllable. The subjects typically reported hearing /da/. More recently, infants have also been found to show the McGurk effect (Rosenblum et al., 1997; Burnham & Dodd, 2004), even though the integration is not always mandatory (Burnham & Dodd, 2004). In this regard, Desjardins and Werker (2004) suggest that in early infancy, independent modes of processing supporting integration are in place, but they are not yet fully specified.

The specific weights attached to auditory and visual (articulatory) cues in infancy are not clear. Auditory cues alone are known to enable subtle speech discrimination in very young infants (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Cheour-Luhtanen, Alho, Kujala, Sainio, Reinikainen, Renlund, Aaltonen, Eerola & Näätänen, 1995; Sambeth, Ruohio, Alku, Huottilainen, & Fellman, 2008). But a key question with regard to auditory-visual integration is under which circumstances visual cues are most informative. One way to address this question is to determine how visual and auditory cues influence learning in phonetic discrimination. Maye, Werker, and Gerken (2002) found that a group comprised of both 6- and 8-month-old infants was influenced by the frequency distribution of consonant-vowel tokens that lay along a /da-/ta/ continuum. They exposed separate groups of infants to unimodal or bimodal frequency distributions of speech sounds from this continuum, and found that the infants exposed to the unimodal distribution failed to discriminate the phoneme contrast along the continuum,

whereas the infants exposed to the bimodal distribution retained their perception of the contrast. Maye, Weiss, and Aslin (2008) conducted a complementary study using pre-voiced/voiced /da-/ta/ and /ga-/ka/ continua, finding that infants exposed to the bimodal distribution showed discrimination of the phonetic contrast, whereas infants exposed to the unimodal distribution and control infants exposed to tones failed to show discrimination. These results suggest that infants are able to use distributional information in auditory speech input to attenuate their pre-existing phonetic discrimination (Maye et al., 2002) or enhance their perception of non-native phonetic contrasts (Maye et al., 2008). Recently, distinct phonetic frequency distributions of this sort were found in natural infant-directed speech. Werker, Pons, Dietrich, Kajikawa, Fais and Amano (2007) found language-specific differences between English and Japanese mothers in the distributions of the cues that they used to distinguish their native-language vowels when teaching words to their infants.

Here we ask whether visual speech cues may also contribute to the distributional learning of phoneme contrasts. We exposed two groups of 6-month-old infants to syllables from the middle range of an auditory /ba-/da/ continuum, using a unimodal auditory frequency distribution, which previously failed to enhance the perception of /da-/ta/ and /ga-/ka/ contrasts (Maye et al., 2002; Maye et al., 2008). Simultaneously, we presented half of the infants with visual speech in a binary visual distribution, using the prototypical visual articulations of the endpoint syllables from the continuum. If these articulations provide infants with a strong cue to the /ba-/da/ contrast despite the unimodal auditory distribution, then they should show evidence of /ba-/da/ discrimination in a post-test where visual cues are absent. The other half of the infants was exposed to the same unimodal auditory continuum, but all speech sounds were combined with the same visual articulation, either /ba/ or /da/. As this unary visual distribution provided a single cue, reinforcing that of the unimodal auditory continuum, we predicted that infants in this group would show no evidence of discriminating the /ba-/da/ contrast.

2. Method

2.1. Participants

A total of forty-eight 6-month-old infants (28 female) of volunteer parents, averaging 6 months 11 days in age (range 5 months 20 days to 6 months 20 days) participated in the experiment. We excluded infants that could not finish the experiment due to crying and infants that sustained a fixed gaze throughout three or more test trials. According to these criteria, five additional infants were excluded from the final sample (three for crying, one for sustained gaze throughout three test trials, and one due to experimenter error). The study was approved by the ethical committee at the School of Psychology of Birkbeck, University of London. One or both parents gave their consent prior to the testing. The infants were assigned to one of two familiarisation conditions (discussed below), with 24 infants (14 female) in each group.

2.2. Apparatus

Infants were seated on their parent's lap 1 m from a 42-in. plasma screen (Sony PFM-42B1E). Auditory stimuli were played through loudspeakers located behind the screen. The sound volume was set to a comfortable listening level. A Macintosh G5 computer using Psychophysics Toolbox for Matlab was used to present the stimuli. Video of the infant's face was recorded with a night vision camera for off-line analyses. The parents wore silencing Peltor headphones playing masking classical music in order to avoid any stimulus-related behaviour that could affect the infants.

2.3. Stimuli

The visual stimuli were videotaped repetitions of a female speaker saying the syllables /ba/ and /da/. One token was selected for each syllable based on the quality of the articulation and the clarity of the soundtrack. In addition, the pitch and intensity of the soundtracks were selected to be approximately equal. The tokens were then trimmed to 1.0 s in duration, the video beginning and ending in a closed-lips position. With lips closed between the visual stimuli, the looped stimulus presentation resulted in a perception of continuously repeating speech syllables.

An auditory continuum was generated to create 21 synthetic syllables, each 200 ms in duration, at equidistant steps between the syllables /ba/ and /da/. Soundtracks of the chosen /ba/ and /da/ video tokens were used as a source for the speech synthesis. See the Appendix for the details of speech synthesis.

Location of the adult /ba/–/da/ phoneme boundary was determined for the continuum in the absence of any visual stimulus. Three English-speaking adults rated ten repetitions of each auditory token in random order as /ba/ or /da/. The phoneme boundary was found to lie at the acoustic midpoint of the continuum (Fig. 1). Eight sounds around

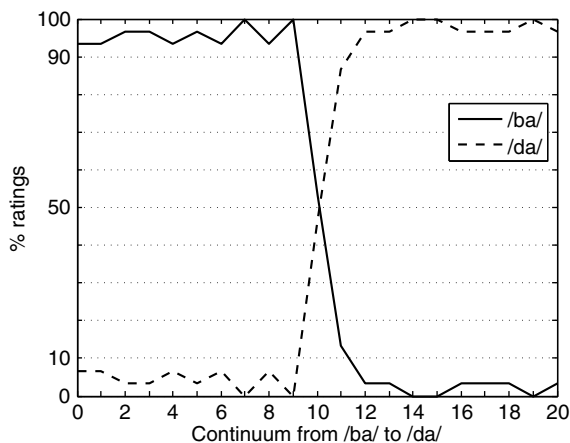


Fig. 1. Average adult ratings ($N = 3$) of the synthesized sounds from /ba/ to /da/ revealed a clear phoneme boundary located at the acoustic midpoint of the continuum. Token 10 was rated approximately equally often as /ba/ and /da/, whereas the other tokens were consistently rated as either /ba/ or /da/.

the midpoint (tokens 3, 5, 7, 9, 11, 13, 15, and 17 in the synthesised continuum) were chosen to comprise the auditory continuum for the study. As the adults primarily perceived the first four of the tokens as /ba/ and the last four tokens as /da/, we refer to the stimuli in the continuum as /ba1/–/ba4/ and /da5/–/da8/.

Three sets of familiarisation stimuli were created. In the familiarisation for the two-category condition, the tokens /ba1/–/ba4/ were dubbed onto a visual articulation of /ba/, and the tokens /da5/–/da8/ onto a visual articulation of /da/. In the two other sets, created for the one-category condition, all tokens were dubbed onto a visual articulation of either /ba/ or /da/. The resulting audio-visual stimuli were 1.0 s long, out of which 200 ms was voiced. The silences between the voiced sections were thus 0.8 s long. The onset of voicing was matched to that on the original soundtrack to achieve auditory-visual synchrony.

Additional stimuli were created for the test phase that did not differ between the one-category and two-category groups. These stimuli included the auditory tokens /ba3/ and /da6/ dubbed onto a static colourful bulls-eye pattern, i.e., the infants did not see visual speech during the test phase.

2.4. Procedure

Infants were assigned to one of two conditions. Infants in each condition were exposed to speech sounds from the earlier defined continuum between /ba/ and /da/. These tokens followed a unimodal frequency distribution (Fig. 3) centered at the average adult category boundary (Fig. 1). In the two-category condition, the visual articulation of a canonical /ba/ or /da/ was presented in synchrony with the auditory token, and this display corresponded to the /ba/ or /da/ side of the midpoint of the continuum. Infants in the one-category group received the same unimodal distribution of speech sounds, but every token for a given infant was paired with either a visual /ba/ ($N = 12$, 7 female) or a visual /da/ ($N = 12$, 7 female) articulation (Fig. 2).

The stimuli presented during the familiarisation phase were organized into six runs of 16 trials each. Each run contained, in a randomised order, one token of /ba1/, /ba2/, /da7/, and /da8/, two tokens of /ba3/ and /da6/, and four tokens of /ba4/ and /da5/ (Fig. 3). Whenever the infants turned their eyes away from the screen, the stimulus presentation was paused and continued once the infants looked back at the screen. The stimulus during which the infant stopped attending was repeated. If the infant did not return to attend to the screen within a few seconds, simple musical samples of xylophone or drum sounds were used to attract the infants' attention. When necessary, swirling geometric shapes were shown on the screen to boost the effect of the musical attractors. Without any pauses, the total length of the familiarisation phase was 2 min 5 s.

After familiarisation, a modified version of the stimulus-alternation preference procedure (Best & Jones, 1998) was used to assess phoneme discrimination. This procedure was identical to that used by Maye et al. (2002). Each infant received eight test trials of two types. During the Repeating test trials, token /ba3/ (every second Repeating

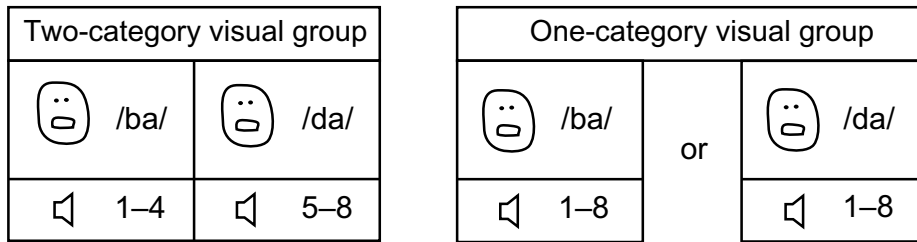


Fig. 2. Two groups of infants received different visual input during familiarisation. In the two-category visual group, two different visual articulations provided information for a two-category representation of the auditory continuum, whereas in the one-category visual group, the visual syllable was always the same for a given infant.

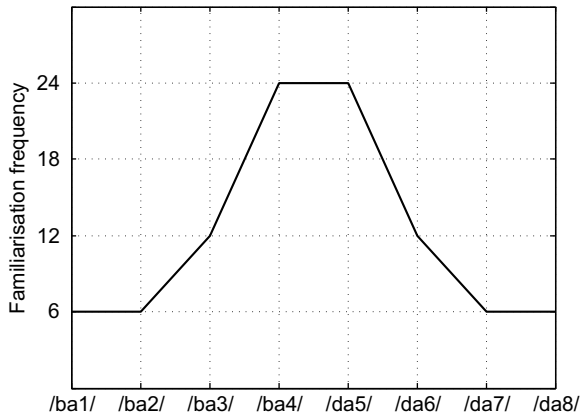


Fig. 3. Unimodal frequency distribution of speech sounds from the /ba-/da/ continuum during familiarisation. Syllables /ba3/ and /da6/ were used in the test trials as the alternating stimuli to test for /ba-/da/ discrimination.

trial) or /da6/ (the rest of the Repeating trials) was repeated 12 times, while in the Alternating test trials tokens /ba3/ and /da6/ alternated. The duration of one test trial was 12.1 s. To reduce the effect of trial order, the order of the test trials was “ABBAABBA”. The first trial type and the starting token within the test trials were counterbalanced across participants.

2.5. Data analysis

The looking times towards the screen in the test trials were scored off-line from the video recordings. The looking behaviour of 16 randomly chosen participants was also coded by a second coder, who was unaware of both the hypothesis of the study and to which group the coded infants belonged. The interrater correlations were .983, and the average absolute difference between the two codings was 0.04 s. The average number of times the infants looked away from the screen during familiarization was scored separately for the two conditions to find any prospective differences in attention between the groups. The numbers (4.75 for infants in the one-category condition and 5.29 for infants in the two-category condition) did not differ significantly ($t[46] = 0.536, p = .594$; two-tailed t -test) suggesting no differences in attention between infants in the two conditions during the familiarisation phase. A mixed-design ANOVA (2 Visual Syllables \times 2 Trial Types) within

the one-category visual group showed no significant interaction of Visual Syllable and Trial Type ($F[1,22] = 0.430, p = .519$) or main effect of Visual Syllable ($F[1,22] = 0.872, p = .361$). Therefore, the data from the infants in the one-category visual groups (/ba/ and /da/) were pooled in the subsequent analyses.

3. Results

Mean looking times for infants in the test trials are provided in Table 1. A mixed-design ANOVA (2 Conditions \times 2 Trial Types \times 2 Trial Orders) showed no significant main effects. The interaction of Trial Type and Trial Order was significant ($F[1,44] = 5.894, p = 0.019$), and the interaction of Condition and Trial Type was marginally significant ($F[1,44] = 3.525, p = 0.067$), but there was no interaction of Condition with Trial Order ($F[1,44] = 0.179, p = .674$), nor any Condition by Trial order interaction ($F[1,44] = 0.945, p = .336$). Planned paired samples t -tests confirmed the hypothesis that the looking times for the two trial types (Alternating vs. Repeating) differed for the infants in the two-category condition ($t[23] = -2.477, p = 0.021$), but not in the one-category condition ($t[23] = 0.336, p = .740$). A non-parametric test confirmed that significantly more infants in the two-category condition (19 of 24) than in the one-category condition (10 of 24) showed longer average looking times for Repeating test trials ($p = 0.017$, two-tailed Fisher’s exact test). The Trial Type by Trial Order interaction was due to significantly longer looking times for the Repeating test trials when the first trial was Repeating ($t[23] = -2.718, p = 0.012$), but no differences in the looking times when the first trial was Alternating ($t[23] = 0.729, p = .474$).

4. Discussion

The results from infants in the one-category visual group reflect the main finding from Maye et al. (2002) by showing that 6-month-olds who listen to a unimodal frequency distribution of tokens from an 8-step synthetic

Table 1
Mean (SE) looking times according to condition and test trial types

	Alternating trials (s)	Repeating trials (s)
One-category condition	7.72 (0.34)	7.64 (0.36)
Two-category condition	7.38 (0.40)	7.83 (0.39)

speech continuum fail to discriminate the end points (and in our case tokens 3 and 6 which are even closer to the category boundary). This was expected since the presence of a uniform (either /ba/ or /da/) visual articulation in synchrony with each auditory token provided further information for the infants that the continuum represented a single category.

More importantly, the results from infants in the two-category group, who were provided with conflicting cues about the number of speech categories (i.e., unimodal auditory distribution = one category; binary visual distribution = two categories) showed evidence of discriminating tokens 3 and 6. That is, despite auditory evidence for a single category, seeing a categorically relevant visual stimulus during exposure enabled infants to discriminate two speech sounds that straddled the category boundary along the auditory continuum. These results suggest that not only did the visual speech affect the discrimination of the heard speech sounds, but this effect survived the loss of the visual articulation information, since the test phase was conducted while infants viewed a neutral bulls-eye display.

Our results extend those of *Maye et al. (2002)* by showing that 6-month-old infants are influenced not only by the frequency distribution of tokens from a synthetic speech continuum, but also that simultaneously presented visual information for speech articulation can influence the learning outcome. When both auditory and visual information were consistent with one category, the exposure resulted in no discrimination of the speech sounds. When there was a conflict between the auditory and visual domains, however, the visual binary information influenced perception of the auditory unimodal distribution, highlighting the phonetic contrast.

While the post-exposure phonetic discrimination of the infants in the two-category visual group was more consistent with the binary visual categorical cues than the unimodal auditory distributional cues, it would be premature to assume that infants of this age would always weight visual cues over the auditory information. It is, for instance, possible that infants have learned the /ba-/da/ contrast from the bimodal auditory distribution present in native-language input during the first postnatal months, forming a natural bias to perceive the two-category representation. Consequently, only when neither auditory nor visual information supports two categories, would the infants fail to discriminate the phonetic contrast.

5. Conclusions

Our results provide evidence of visual effects on phoneme discrimination and learning in infants. They not only show that visual information about speech articulation enhances phoneme discrimination, but more importantly that this enhancement during a learning phase carries over into an auditory-only post-test. Thus, synchronous visual speech can influence the statistical information embedded in auditory speech, and may contribute to the learning of phoneme boundaries in infancy.

Acknowledgements

We thank Sarah Lloyd-Fox, Tamsin Osborne, Agnes Voilein, Leslie Tucker, Hanife Halit, Jane Singer, as well as the two anonymous reviewers for their valuable contributions. This research was supported by the Finnish Cultural Foundation, James S. McDonnell Foundation, and Signe and Ane Gyllenberg Foundation.

Appendix A. Speech synthesis

Natural speech tokens of /ba/ and /da/, extracted from a video soundtrack, were used as a source for the speech synthesis. Both of these syllables consist of a voiced plosive followed by a vowel. Therefore, their waveforms are known to comprise a short noise burst followed by a clearly longer voiced segment (*Kent & Read, 1992*). The voiced segment begins with formant transitions that lead to a steady-state vowel, with a slight drop in fundamental frequency across the entire syllable.

The stimulus generation was started by measuring the durations of the different sections from the natural syllables, which were then used in all the sounds involved in the stimulus continuum. The waveforms of the noise bursts in the natural /ba/ and /da/ sounds were then separated with the help of linear predictive (LP) analysis (*Makhoul, 1975*) into an excitation and a filter. The noise burst waveform of each sound in the continuum was obtained by interpolating the LP excitation and filter computed from /ba/ and /da/. The LP excitation was interpolated by using time-domain cross-fading from the waveform computed from /ba/ into the waveform computed from /da/. The LP filter was processed by first expressing it in terms of the line spectral pairs (LSPs) (*Itakura, 1975*) and then by interpolating these values on the Bark scale (*Moore, 1982*) between the corresponding values obtained from /ba/ and /da/. For the voiced segment of the stimuli, the waveforms of the natural sounds /ba/ and /da/ were first separated into the glottal flow and vocal tract filter by using the Semi-synthetic Speech Generation (SSG) technique presented in *Alku, Tiitinen, and Nääätänen (1999)*. This computation was performed separately for the segments over the plosives /b/ and /d/ and the vowel /a/.

The vocal tract filter computed with SSG was expressed in the form of second order all-pole sections so that the formant centre frequencies and bandwidth could be interpolated on the Bark scale between those extracted from /b/ and /d/ to obtain the corresponding intervening values used in the continuum. The waveform of the voiced section for each stimulus in the continuum was produced by exciting the interpolated filters with a synthetic glottal flow waveform generated with the Liljencrants-Fant pulseform (*Fant, 1993*). Finally, the sound waveform of each auditory stimulus was obtained by concatenating the waveform representing the noise burst into the corresponding waveform of the voiced section and by normalizing the sound intensity by adjusting the digital energy of each waveform to the same value.

References

- Alku, P., Tiitinen, H., & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology*, *110*, 1329–1333.
- Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior & Development*, *21*, 295.
- Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*, 204–220.
- Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., et al (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*, 53–58.
- Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*, 187–203.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 431–461.
- Fant, G. (1993). Some problems in voice source analysis. *Speech Communication*, *13*, 7–22.
- Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics*, *45*, 34–42.
- Itakura, F. (1975). Line spectrum representation of linear predictive coefficients of speech signals. *Journal of the Acoustic Society of America*, *57*(Suppl. 1), 35.
- Kent, R., & Read, C. (1992). *The acoustic analysis of speech*. San Diego, USA: Singular Publishing Group.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.
- Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.
- Makhoul, J. (1975). Linear prediction: A tutorial review. *Proceedings of the IEEE*, *63*, 561–580.
- Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, *55*, 1777–1778.
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge: MIT Press MA.
- Maye, J., Weiss, D., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*, 122–134.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Moore, B. (1982). *An introduction to the psychology of hearing*. London UK: Academic Press.
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behaviour & Development*, *22*, 237–247.
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*, 191–196.
- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*, 347–357.
- Sambeth, A., Ruohio, K., Alku, P., Huotilainen, V., & Fellman, M. (2008). Sleeping newborns extract prosody from continuous speech. *Clinical Neurophysiology*, *119*, 332–341.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Amano, L., & Fais, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, *103*, 147–162.