INFANCY, *3*(2), 249–259

BRIEF REPORTS

Testing Infants' Discrimination With the Orientation Latency Procedure

Ariel Gout, Anne Christophe, and Emmanuel Dupoux
Laboratoire de Sciences Cognitives et Psycholinguistique
CNRS-EHESS, Paris

A new discrimination procedure based on the measurement of visual orientation latency to speech stimuli is introduced. Each participant listens to a series of short familiarization test trials. In each trial, 5 to 7 centrally-presented familiarization stimuli are followed by laterally-presented test stimuli. Infants were found to orient faster to different-category than to same-category test stimuli. This result was found despite a high degree of prosodic variability in the familiarization and test stimuli introduced by changes in talker and speaking rate. The combination of a multitrial design with use of acoustic and prosodic variability seems suitable for studying the representation of phonological categories.

In this article, we introduce a new discrimination procedure designed to explore the representation of phonological categories in infants. Quite a lot of literature is available on the development of phonetic and phonemic categories (see, e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) but very little on other types of phonological categories, such as syllabic structure or stress. Yet studies with adults tell us that the native phonology influences the perception of these properties, just as it influences phonetic and phonemic categorization (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Sebastian-Gallés, & Mehler, 1997). It is therefore important to uncover the developmental course of the perception of these phonological categories.

Requests for reprints should be addressed to Anne Christophe, LSCP, MSH, 54 Bd Raspail, 75270 Paris Cedex 06 France. E-mail: anne@lscp.ehess.fr

Studies with adults have shown that people whose native language does not exploit stress or some syllable types cannot perform well in simple discrimination tasks, at least when stimuli display acoustic–prosodic variability (Dupoux et al., 1999; Dupoux et al., 1997). By contrast, in the absence of variability, adults generally perform well irrespective of their native language phonology. For instance, stress is not a relevant phonological category in French (French words never differ solely on stress), but it is one in Spanish. Dupoux et al. (1997) observed that in a same–different task using trisyllabic stimuli differing solely in stress position, French participants performed as well as Spanish participants when identical tokens were presented in the "same" condition. By contrast, when several speakers and acoustically varied stimuli were presented, French participants performed significantly worse than Spanish participants. These results suggest that participants relied on an acoustic level of representation when identical tokens were compared, but that acoustic–prosodic variability prevented them from doing so, and forced them to represent the stimuli categorically and to use a more abstract level of representation. Similarly, for the case of syllabic types, Japanese uses mainly a CV syllabic structure: Stimuli like *igmo* are illegal in Japanese. Japanese adults were shown to perceive an illusory *u* vowel in stimuli like *igmo* (Dupoux et al., 1999). They were unable to consistently distinguish between *igmo* and *igumo* in a same–different task, at least when several speakers produced the stimuli (Dehaene-Lambertz, Dupoux, & Gout, 2000). These adult experiments show that it is crucial to include speaker and acoustic–prosodic variability when studying the representation of phonological categories like syllabic structure or stress.

When studying the development of these phonological categories in infants, it is therefore important to use a technique in which many acoustically different tokens of each category are presented. However, acoustic–prosodic variability has sometimes been shown to drown out the relevant dimension and confuse infants. For instance, Bertoncini and her colleagues, using a non-nutritive sucking habituation–dishabituation paradigm, observed that newborn infants failed to discriminate two- versus three-syllabic nonwords in the presence of prosodic variability (Bertoncini, 1993) but succeeded in its absence (Bertoncini, Floccia, Nazzi, & Mehler, 1995). It is important to notice that in this paradigm there is only one familiarization phase with many different tokens from one category. These results suggest that when infants listen to many tokens exemplifying one single category in the habituation or familiarization phase, they may fail to notice the dimension that will become relevant in the comparison with the test stimuli (here the number of syllables; see also Jusczyk, Pisoni, & Mullenix, 1992, for the disruptive effects of speaker variability on phonetic discrimination in a similar procedure). One notable exception to this pattern is the series of experiments by Kuhl (1983), who showed with the conditioned head-turn procedure that infants could successfully be trained to ignore (synthetic) speaker variation when categorizing vowels.

We present a new experimental procedure mimicking the adult procedures that successfully showed cross-linguistic differences in the processing of phonological categories such as stress and syllabic structure. The experimental design contains many short familiarization-test trials, in which infants hear several speakers in each familiarization phase, and a speaker change (from female to male) between familiarization and test in each trial. These changes are constant across trials and therefore become less salient over time. Infants may then learn to ignore the speaker variability and focus their attention on the varying phonological dimension that is relevant to the test.

More precisely, the discrimination procedure we introduce is based on the measurement of visual orientation latency to speech stimuli, combined with a multitrial within-subjects design. We define orientation latency as the time an infant needs to initiate an ocular saccade in the direction of a lateral loudspeaker from which the stimuli are presented. The measurement of visual orientation latency has already been used successfully in language categorization experiments: 2- and 4-month-old infants have been shown to orient significantly faster in response to maternal sentences than to foreign sentences (Bosch & Sebastian-Gallés, 1997; Dehaene-Lambertz & Houston, 1998). Up to now, orientation latency has been measured for stimuli presented in isolation, and it reflected infants' intrinsic interest for each category of stimuli (e.g., sentences in the mother tongue vs. sentences in a foreign language). As such, it could not be used to study the discrimination between two categories of stimuli that do not differ in attractiveness (e.g., two foreign languages). To study discrimination, we combined the measurement of orientation latency with a specific experimental design, borrowed from evoked related potential (ERP) studies (as in Dehaene-Lambertz & Dehaene, 1994). This ERP design consists of many short familiarization-test trials; each trial starts with four familiarization stimuli that all belong to the same category and continues with "test" stimuli that either belong to a different category (experimental trials) or to the same category (control trials) as the familiarization stimuli. The new combination of this multitrial familiarization-test experimental design and the measurement of orientation latency allows us to compare orientation latencies to the same stimulus, depending on what precedes it (stimuli from either the same or a different category).

To evaluate this new procedure, we used stimuli that have been validated in a cross-linguistic adult experiment about syllabic structure. Dehaene-Lambertz et al. (2000) performed an ERP experiment in which each trial contained four familiarization stimuli pronounced by several female voices, followed by one test stimulus pronounced by a male voice. Adult participants had to judge whether familiarization and test stimuli were "same" or "different" by pressing two response keys. Japanese adults failed to respond "different" to *igumo–igmo* trials (only 20% different responses), in sharp contrast with French adults (95% different responses). This highly significant difference in behavior between French and

Japanese adults suggests that these stimuli, presented in this task (A…AX), allow us to tap a phonological rather than acoustical level of representation. For this reason we used the same stimuli, and an almost identical task, with infants, using orientation latency as a measure.

To sum up, we wish to assess whether this new procedure is adequate to study the development of phonological categories such as syllabic structure or stress: It consists of presenting many acoustically different tokens for each category, preventing participants from relying on an acoustic representation to perform the task. In addition, all experimental conditions are presented to each infant (a design that entails within-subjects statistics).

Method

In each trial, familiarization speech stimuli were presented centrally while a dynamic, colored stimulus was presented on a screen situated in front of the infant to keep his or her attention centered. Test stimuli were then presented laterally, randomly to the left or right of the infant, and the center screen was darkened: Infants were expected to orient toward the lateral loudspeakers on each trial. In half of the trials, test stimuli were from the same category as familiarization stimuli (control trials), and in the other half they were from a different category (experimental trials). Trials lasted approximately 10 sec each; thus each infant could listen to many different trials while keeping the experiment short (5–15 min for 25–75 trials). This allowed the experimental design to be fully within-subjects because each infant was exposed both to control and experimental trials (including counterbalancing, e.g., *igumo–igmo* vs. *igmo–igumo*).

Stimuli. The stimuli were five item pairs: *igmo/igumo, igna/iguna, ikno/ikuno, okna/okuna, and ogma/oguma* (same stimuli as in Dehaene-Lambertz et al., 2000). The familiarization stimuli and the test stimuli were constructed as follows:

> The familiarization stimuli set consisted of 60 different items (six tokens of each member of five pairs) produced by 6 female Japanese speakers. The stimuli were digitized at 16 kHz/16 bits on an OROS AU22 board. Many trisyllabic stimuli (*igumo*-type) were naturally produced; 60 of them were then selected by French and Japanese listeners so that all items were acceptable in both languages. Because Japanese speakers couldn't help inserting a short vowel between the consonants of the *igmo*-type stimuli, the remaining vocalic glottal pulses were manually removed in a waveform editor.
>
> The test stimuli set consisted of 10 synthetic items (five pairs). They were synthesized with the MBROLA speech synthesizer (Dutoit, Pagel, Pierret, Bataille, & van der Vrecken, 1996), using as a model one token each of *igmo* and *igumo*, naturally produced by a male Japanese speaker. The phonemes of the test stimuli

had the same duration and pitch as the natural model. Phoneme duration and pitch were also constant across the five pairs. The third phoneme (*u* or the second consonant of the cluster) always started 240 msec after the onset of the file. This feature of the stimuli may be useful, because it is possible that gaze latency is influenced by the durational properties of the stimulus. Mean stimulus duration was 700 msec.

Thus there was an important speaker variability both within the familiarization phase (6 female speakers), and between familiarization and test (from female to male).

Experimental design. The experimental design consisted of five conditions: two same (*igumo–igumo* and *igmo–igmo*), two different (*igumo–igmo* and *igmo–igumo*), and one distractor condition (see Table 1). In this latter condition, the vowel *u* was changed to *i* (*igumo* to *igimo*). There were 20 experimental and control trials (each stimulus pair appeared in all four experimental conditions) and 5 distractor trials (containing the vowel change). Trials were grouped into three blocks of 25. Within each block, trial order was random and different for each infant. Each block lasted approximately 5 min; the experiment continued until the infant stopped attending to the stimuli, with a minimum duration of one block (25 trials) and a maximum duration of three blocks (75 trials = 15 min).

Apparatus. The experiment was conducted in a soundproof room. The infant was seated in a high chair in the center of the room, with one parent seated behind the chair, listening to masking noise[1] through sound-attenuating headphones. The infant faced a computer screen, above which was mounted a video camera. Loudspeakers associated with flashing lights were mounted at eye level, on either side of the infant. A dark curtain suspended around the room shielded all apparatus but the computer screen and flashing lights. An amplifier, two computers, a response box, a video recorder, and a TV monitor were located outside of the testing room. The experimenter watched the infant on the TV monitor and coded the infant's gaze direction by pressing the response box buttons (center, left, right). The experimenter was unable to hear the auditory stimuli and was thus blind to the condition of each trial. The entire experimental session was recorded on a Sony SVO 96-20 video recorder equipped with a time code and special board for communication with the computer. Stimulus presentation, randomization, and behavioral response measurement (via experimenter coding) were done using the EXPE software package (Pallier, Dupoux, & Jeannin, 1997) on a PC-compatible computer with a Proaudio Spectrum 16 D/A board.

Procedure. Each trial began with a familiarization phase: The central screen displayed colorful, dynamic spirals while familiarization stimuli were

[1]The masking noise was made using several superimposed streams of continuous speech. It sounds like cocktail party noise and is very efficient to mask speech stimuli.

TABLE 1
Experimental Design

Experimental Condition	*Familiarization (5 to 7 Stimuli)*	*Test (× 3)*
Different	*igumo*	*igmo*
Same	*igmo*	*igmo*
Different	*igmo*	*igumo*
Same	*igumo*	*igumo*

played centrally (through both lateral loudspeakers simultaneously). Stimuli were presented in semirandom order with an interstimulus interval (ISI) of 500 msec. During this phase, the infant generally stared at the central screen. The number of familiarization stimuli varied randomly from five to seven to prevent infants from making anticipatory head turns after hearing the last familiarization stimulus. The test phase began as soon as the familiarization phase ended: The central screen was then darkened, and one of the lateral loudspeakers delivered the test item three times in succession, again with an ISI of 500 msec. This repetition ensured that the infants had sufficient time to make an orientation response. The side of delivery was determined at random. In most trials, the infant turned his or her head toward the sound. As soon as the infant made a head turn of at least 30° in either direction, the experimenter pressed the corresponding button. If the head turn was in the direction of the sound, as determined by the computer, the corresponding side light began to flash and continued flashing until the trial's completion (this was included to maintain infants' spontaneous orientation response throughout the experiment). The test phase ended 2 sec after the third repetition of the test item; the overall duration of the test phase was 5.8 sec on average. The light stopped flashing at the end of the test phase, and the next trial began immediately.

Coding. Orientation latency was defined as the time infants needed to initiate a saccade from the central screen to the lateral active loudspeaker. Coding was offline and semiautomated: The video recorder was controlled by the computer (using the EXPE software package). The computer fast-forwarded to the first frame of the test phase so that the coder could check whether the infant was looking at the central screen and his or her eyes were clearly visible. If this was not the case, the trial was considered invalid because a proper latency could not be computed when the starting point of an ocular saccade was not central. Whenever infants' eyes were centered on the first frame of the test trial, the coder determined the first frame at which infants' eyes could be seen to have shifted from the center. In most cases, this frame corresponded to the beginning of a clear saccade toward one of the lateral loudspeakers. In a few instances, however, infants shifted their eyes slightly to either side (or downward) before launching a real saccade: In these cases, the onset of the real saccade was coded. Orientation latency was computed automatically as the difference between the first frame of the ocular saccade and the first frame of the test phase (each frame lasted 40 msec). The coder then entered the side of the head turn (orientation mistakes were identified by comparing this with the side of presentation of the test stimuli). Occasionally a trial could not be coded, for example, because no head turn occurred or the infant cried. The coding data were saved by the computer and used in the analyses. The coding procedure lasted about 15 to 20 min for each participant. All infants were coded by two independent coders.

Participants. Twenty-four infants participated in this experiment, so they had at least two trials in each of the four experimental conditions (*igumo–igmo*; *igmo–igumo*; *igumo–igumo*; *igmo–igmo*). Their mean age was 10.6 months (range = 9.8–11.1 mo). In addition, 15 infants were tested, but their data were not kept in the analysis for the following reasons: technical problem (3), less than two valid trials in at least one condition (11), or more than 40% orientation mistakes throughout the experiment (1).

Results

Coding reliability. The overall reliability coefficient was .82 (i.e., the correlation coefficient between the latencies computed by both coders). A more detailed analysis revealed that out of 1,282 trials overall, coders agreed whether or not a trial was codable on 93.7% of trials (61.4% of trials were judged codable by both coders; 32.3% of trials were judged noncodable by both coders, mainly because the infants' eyes were not centered on the first frame of the test phase, and 6.3% of trials were judged codable by one coder and not by the other). For trials judged codable by both coders, both latencies differed by more than two frames in only 2.8% of all trials (when both latencies were identical, this value was retained; whenever they differed by one or two frames, the mean latency was used in the analyses). All cases of discrepancy (6.3% + 2.8%) were reviewed by both coders until they reached an agreement. The mean number of valid trials per infant was 27.8 (range = 12–49 across all conditions).

Latency analysis. Only head turns toward the correct loudspeaker were analyzed. Orientation errors represented 5.7% of the data. Latencies below 260 msec represented 3.6% of the data, and about half of them were orientation mistakes: Therefore these movements were considered anticipatory responses and were not included in the analyses. Latencies over 3 sec (the duration of the test stimuli) were considered outliers and were assigned the cutoff value of 3,000 msec (these

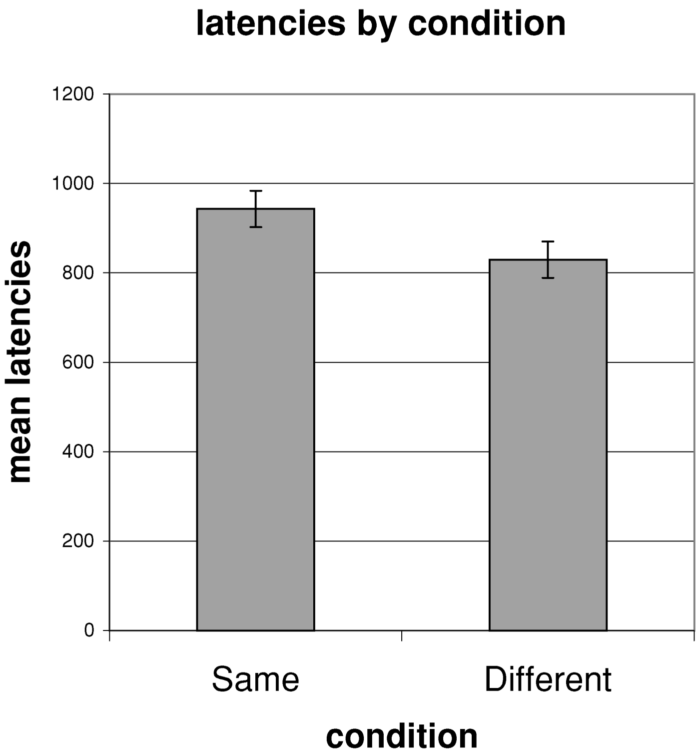

FIGURE 1 Orientation latencies for the same and different conditions; error bars represent the standard error of the mean of the difference between conditions.

latencies represented only 1.7% of the data). Overall mean latency was less than 1 sec; analyses without cutoffs gave the same results.

An analysis of variance (ANOVA) with participants as the random variable was performed on mean latencies by condition. There were two within-subjects factors: the experimental factor, same/different, and a counterbalancing factor, test item (whether the test item contained a medial vowel or not). There was a main effect of the experimental factor, with different trials generating faster latencies than same trials (829.5 vs. 943 msec, effect size 113.5 msec), $F(1, 23) = 7.8$, $p < .01$. The counterbalancing factor had no effect and did not interact with the experimental factor (both $F < 1$). A nonparametric analysis also showed a significant difference between the same and different conditions (Wilcoxon Signed Ranks test, $Z = 2.5$, $p < .02$; 17 infants showed greater latencies in the same condition, 7 infants showed the reverse pattern). There was no indication that the response pattern changed over the course of the experiment: An analysis restricted to the first block of the experiment (first 25 trials) revealed an advantage for different over same of the same amplitude as in the overall analysis (different, 728 msec; same, 860 msec; effect size 132 msec), $F(1, 23) = 7.9$, $p < .01$.[2]

Discussion

This experiment demonstrates that infants initiate an ocular saccade significantly faster toward a lateral loudspeaker when there is a mismatch between the test stimuli presented laterally and familiarization stimuli presented centrally. More specifically, this experiment shows that infants perceive a difference between stimuli like *igmo* and *igumo*. These stimuli differ only in the presence or absence of a medial vowel, which results in a different number of syllables (two vs. three). In addition, in our experiment they were pronounced by different speakers, a feature that ensures that adult participants cannot rely on an acoustical level of representation to compare them (as shown by a cross-linguistic adult experiment, see introduction). Infants apparently had no difficulty in ignoring the speaker variability: The presence of many short familiarization-test trials may help infants to focus their attention on the relevant test dimension (in addition, learning to ignore speaker variability probably happens fast, within a few trials, because there is no difference between the first block of 25 trials and the remainder of the experiment). These results indicate that this design is adequate to study infants' discrimination of stimuli that exhibit an important within-category acoustic–prosodic variability, in conditions in which adults are shown to rely on a language-specific phonological representation. It is therefore plausible that this experimental technique is adequate to study language-specific phonological representations in infants (even though this conclusion will be fully supported only by cross-linguistic experiments).

Another advantage of having many trials is that the experimental design was fully within-subjects and used several different experimental items: Each infant was exposed to several experimental and control conditions, represented by several different items, and counterbalanced within participants. Because of this, all factors in the ANOVA were within-subjects factors, which typically makes the analyses more powerful than when between-subject factors are used. Nevertheless, the orientation latency procedure does not provide individual results, in the sense that every single participant would show the effect in the right direction: Even though most infants oriented faster toward different than same stimuli, 7 infants out of 24 showed results in the reverse direction (this observation is similar

[2]An analysis of looking times toward the lateral loudspeakers as measured online during the experiment revealed no effect whatsoever: same, 3.28 sec ($SE = 0.09$ sec); different, 3.37 sec ($SE = 0.15$ sec), $F(1, 23) < 1$; looking times are necessarily short because the test phase for each trial lasted only 5.8 sec. The vowel change condition (*igumo/igimo*) generated latencies of 852 msec, similar to the different condition, even though the difference between vowel change and same conditions did not reach significance, $F(1, 23) = 2.2$, $p = .15$ (two-tailed).

to what is typically obtained with other behavioral techniques available with infants of the same age, such as the Head-Turn Preference Procedure).

We chose to evaluate the procedure with 11-month-old infants because many phonological properties of the native language appear to be mastered by this age. This age range seems appropriate to conduct cross-linguistic studies in the hope of observing language-specific behavior. Given the nature of the response, the technique should in principle be appropriate to test infants from a wide age range, possibly between 6 months and 18 months or older (when clear orientation responses can be obtained), although the actual experiments remain to be done.

What are the possible interpretations for infants' faster orientation latencies in the different than in the same condition? One interpretation is that infants find the novel stimuli more interesting and orient faster in response to interesting stimuli. This result may be seen as congruent with the language categorization experiments relying on orientation latency, where it was observed that infants orient faster toward their mother tongue (interesting) than toward a foreign language (uninteresting). The language categorization experiments can receive an alternative explanation: Foreign language sentences generated an increased processing cost, which interfered with the orientation mechanism and slowed down responses in this condition. This explanation in terms of interference cannot handle the results reported here: Indeed, there is no reason whatsoever why the same condition should be harder to process than the different condition. Additional experiments are necessary to determine if the difference in orientation latency is always in the same direction, irrespective of the age of the infants and the nature of the stimuli involved. Potentially, a study coupling electrophysiological and behavioral measurements may offer some information about the neuronal mechanisms involved during this kind of task.

Conclusion

The results of this experiment demonstrate that orientation latency, when used with a series of short familiarization-test trials, is suitable to study discrimination. This procedure seems particularly well suited to study the development of phonological representations in infants of approximately 1 year of age because it works well in the face of substantial acoustic–prosodic variability. It therefore allows experimenters to study the representation of abstract phonological dimensions such as stress, tone, or syllabic structure.

ACKNOWLEDGMENTS

The work reported in this article was made possible by a poste d'accueil INSERM to Ariel Gout, as well as a grant from the Convention CNRS GIS 97/C/28, the

Convention CNRS GIS 99/N/35, and the Action Concertée Incitative Cognitique 00/N60/0447/1. Infant testing was performed at the Maternité Port-Royal, Paris. We would like to thank Sylvie Margules for recruiting infants, and Ghislaine Dehaene-Lambertz, Katherine White, Richard Aslin, and two anonymous reviewers for useful comments on this article.

REFERENCES

Bertoncini, J. (1993). Infants' perception of speech units, primary representation capacities. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 249–257). Dordrecht, the Netherlands: Kluwer.

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech, 38*, 311–329.

Bosch, L., & Sebastian-Gallés, N. (1997). Native-language recognition abilities in four-month-old infants from monolingual and bilingual environments. *Cognition, 65*, 33–69.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*, 292–295.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*, 635–647.

Dehaene-Lambertz, G., & Houston, D. (1998). Faster orientation latencies toward native language in two-month-old infants. *Language and Speech, 41*, 21–43.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1568–1578.

Dupoux, E., Pallier, C., Sebastian-Gallés, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language, 36*, 406–421.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vrecken, O. (1996). *The MBROLA Project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes*. Paper presented at the ICSLP'96, Philadelphia. Available: http://tcts.fpms.ac.be/synthesis/mbrola.html

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*, 303–306.

Jusczyk, P. W., Pisoni, D. B., & Mullenix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition, 43*, 253–291.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*, 263–285.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Pallier, C., Dupoux, E., & Jeannin, X. (1997). EXPE: An expandable programming language for online psychological experiments. *Behavior Research Methods, Instruments, & Computers, 29*, 322–327.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.