Developmental Science 3:3 (2000), pp 333–343

PAPER

Unfamiliar voice discrimination for short stimuli in newborns

Caroline Floccia,[1] Thierry Nazzi[2] and Josiane Bertoncini[3]

1. *Laboratoire de Psychologie, Université de Franche-Comté, Besançon, France*
2. *Department of Psychology, University of California, Berkeley, USA*
3. *Laboratoire Cognition et Développement, CNRS, Paris V, France*

Abstract

In order to determine the minimal information required for newborns to discriminate unfamiliar voices two experiments were performed using the presentation of single disyllabic words uttered by male and female speakers. In the first experiment, utilizing the standard high-amplitude-sucking procedure, no significant discrimination was obtained. Hypothesizing that this failure in discrimination could be due to a deficient attention at the unique moment of stimulus change, a second experiment was performed in which the same to-be-discriminated stimuli alternated every minute and in which multiple tokens of the same word were presented to increase stimulus variability. Evidence for voice discrimination was obtained, suggesting that newborns are able to characterize unfamiliar voices on the basis of restrained vocal-tract related information and minor prosodic information. Remarks addressing the consequences of stimulus organization upon attentional demands in experimental procedures in young infants are also presented.

Analyses of a variety of different speakers' production reveal that many cues are available for voice distinction. In addition to vocal-tract related parameters such as $F0$ and timbre (e.g. Fant, 1973), voices also differ on phonetic and prosodic parameters (intonational variations including those of $F0$, rhythmic cues and intensity variations) (e.g. Endres, Bambach & Flösser, 1971; Abberton & Fourcin, 1978). In adults the discrimination and recognition of voices requires relatively little information – a single syllable will suffice. Even in adult rhesus monkeys studies have shown that, after training on complete sentences, they are able to recognize a speaker's voice using only isolated syllables (Shen, Lin & Wu, 1992).

Young infants' discrimination of human voices has been demonstrated several times, most usually in the recognition of the maternal voice (Mehler, Bertoncini, Barrière & Jassik-Gerschenfeld, 1978; DeCasper & Fifer, 1980; Spence & Freeman, 1996), or in the discrimination of unfamiliar voices (DeCasper & Prescott, 1984; Spence & Freeman, 1996; see also Lecanuet, Granier-Deferre, Jacquet, Capponi and Ledru (1993) for a study with fœtuses). All of these studies were conducted using sentence-like stimuli uttered in an infant-directed speech mode. The aim of the present study is first to determine whether newborn infants are able to distinguish two unfamiliar voices on the basis of minimal information carried by a single disyllabic word. However, because this short stimulus offers far less variability and information than sentences on multiple aspects known to be effective in attracting attention, we were aware of the potential attentional 'problems' related to the use of such stimuli. Therefore focus was placed upon the way the stimuli were presented or rearranged in order to enhance neonates' attention upon voice differences. For that purpose, two experimental procedures were compared in Experiments 1 and 2.

We will begin by presenting a brief review of previous studies of voice processing in early infancy to show the kinds of information required for voice discrimination.

Early preference for the maternal voice over a stranger's has been demonstrated when the speakers are producing sentences or pieces of spontaneous speech in an infant-directed speech mode (Mehler *et al.*, 1978; DeCasper & Fifer, 1980; Spence & Freeman, 1996). This preference can still be obtained when mother's and stranger's speech is low-pass filtered (Spence & Freeman, 1996).

Address for correspondence: Caroline Floccia, Laboratoire de Psychologie, Université de Franche-Comté, 30 rue Mégevand, 25000 Besançon, France; e-mail: caroline.floccia@univ-fcomte.fr

However, it has also been observed that, for two cases, no maternal voice preference can be found. The first case concerns whispered sentences (Experiment 2 of Spence & Freeman, 1996): whispering is a speech manner that removes all the voicing information ($F0$ and timbre), and thus intonation variations, while keeping most rhythmic cues intact. The second was demonstrated with backward read sentences (Mehler *et al.*, 1978) in which natural sentence prosody is disturbed. Taken together, these data suggest that the main information upon which infants base their maternal voice recognition is of a prosodic nature.

Discrimination of unfamiliar voices has also been demonstrated using sentence-like stimuli in both newborns (DeCasper & Prescott, 1984; Spence & Freeman, 1996) and fœtuses (Lecanuet *et al.*, 1993). Although newborns do not show any preference for their father's voice in the days following birth, DeCasper and Prescott (1984) obtained discrimination of two male voices uttering various sentences. Using a single sentence stimuli Lecanuet *et al.* (1993) reported a significant cardiac deceleration in the 36–39 week fœtuses at the moment of voice change. Such discrimination has also been demonstrated in newborns with phonetically varied whispered sentences (Experiment 4 of Spence & Freeman, 1996).

In summary, it appears that, for maternal voice recognition or unfamiliar voice discrimination, prosodic information such as rhythm, intensity variations and intonation variations all need to be kept intact in the stimuli.

As mentioned in the introduction, all of the studies presented thus far have used sentence-like utterances as stimuli. In such complex and rich stimuli, many different kinds and levels of information are available to allow production characterization. These include vocal-tract related cues, such as $F0$ value and timbre, but also speaker-specific prosodic pattern (an extreme example would be two productions uttered with two different regional accents) and phonetic information (speaker-specificity in pronouncing vowels or consonants for instance).[1] In addition, it is known that all these types of acoustic information carried by speech are highly interwoven. As noted by Crowder (1993, p. 139): 'To the extent that target voices carry idiosyncratic accents, speech impediments, or the like, subjects can use these non-vocal characteristics, rather than vocal timbre itself, for recognition.'

One of the conclusions that can be drawn from these previous studies on early voice discrimination capacities is that infants extract from sentence-like stimuli some speaker-specific prosodic pattern (see Goggin, Thompson, Strube & Simental, 1991) that allows voice discrimination or recognition. This pattern probably entails use of vocal and non-vocal information, to use Crowder's (1993) terminology, or, to put it differently, vocal-tract related cues and prosodic patterns.

The question we would like to ask concerns utterance length, i.e. how long should utterances be to allow infants to characterize speaker-specific prosodic pattern? Is the presentation of rich sentence-like stimuli really required, or would infants still discriminate voice changes if only single words were presented? In such a short and phonetically identical context there is far less information about voices, in so much that this difference is quantitative, not qualitative. Indeed speaker-specific prosodic variations based upon long duration, as in sentences, are considerably reduced. Vocal-tract related cues, even if they are still available, have less variation space, and might lead to an under-specification of voice quality.

In addition to the reduction of information in the presentation of single words, the relative poverty of variation between single words and sentences might have a secondary effect, i.e. a reduction of sustained attention during the experiment that might impede discrimination. This problem would not have been encountered in previous studies in which sentence-like stimuli were used whose prosodic variations are well known to capture infants' attention. This is shown by early preference for motherese (e.g. Fernald, 1985), or by the essential role prosody plays in early language acquisition, as was demonstrated in natural language discrimination (e.g. Nazzi, Bertoncini & Mehler, 1998), word segmentation or syntactic parsing (see a review by Morgan, 1996; Jusczyk, 1997).

However, voice discrimination using single word stimuli should still be possible in newborns in that other studies with short non-speech sequences report that infants can discriminate changes in mean $F0$ value (infants aged at least 1 month: Wormith, Pankhurst & Moffitt, 1975; Trehub, Thorpe & Morrongiello, 1985; Karzon & Nicholas, 1989) or timbre changes (Best, Hoffman & Glanville, 1982; Bundy, Colombo & Singer, 1982; Clarkson, Clifton & Perris, 1988; Bertoncini, Morais, Bijeljac-Babic, McAdams, Peretz & Mehler, 1989; Trehub, Endman & Thorpe, 1990), with the restriction that non-speech stimuli are less complex than speech stimuli.

[1] Although adults have been shown to use phonetic elements to distinguish speakers (see Remez, Fellowes & Rubin, 1997), the results obtained with whispered speech in newborns (Experiment 2 of Spence & Freeman, 1996) and backward read utterances (Mehler *et al.*, 1978) suggest that fine-grained phonetic information may not be available for infants to perform speaker discrimination. Therefore we will not question the role of this information in early voice discrimination.

The first experiment tested whether newborns can discriminate two unfamiliar voices uttering one single word by using the standard discrimination procedure available with newborn infants, i.e. the high-amplitude-sucking (HAS) procedure.

Experiment 1

We tested discrimination between a female token of the French word /nuga/ and a male token of the same word, using a standard HAS procedure. This procedure typically consists of the presentation of short speech sounds contingent upon infants' HAS responses. After a familiarization period with the first stimulus, terminating when the infant reached a sucking rate decrease criterion, a testing phase took place that lasted 4 min. For the experimental group stimulation was shifted from one voice to the other, while the speaker remained unchanged for the control group. Greater increases in HAS rates after the change in the experimental group compared with the control group would indicate that infants discriminated the stimulus change.

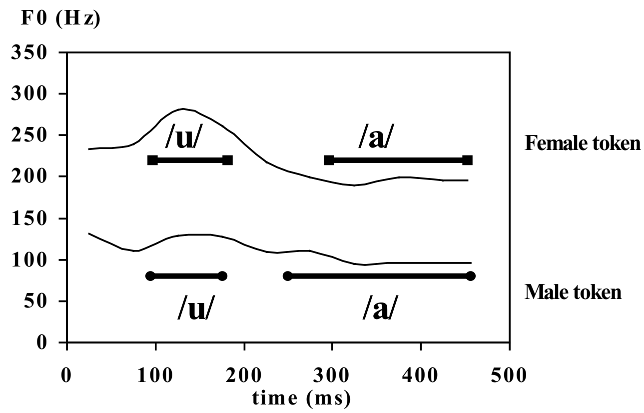


Figure 1 *Pitch contour of the female and male tokens of the word /nuga/ presented in Experiment 1. Duration of vowels /u/ and /a/ are reported for each token.*

Stimuli

A French disyllabic word, /nuga/, was extracted from a list of phonetically varied isolated words uttered with steady intonation, from a native female and male speaker. The duration of the female token was very similar to that of the male token (452 ms and 456 ms respectively). Pitch varied between 180 and 300 Hz for the female token, and between 90 and 150 Hz for the male (see Figure 1). The stimuli were registered on two 16 min tapes (one for the female speaker, one for the male), with a stimulus onset asynchrony (SOA) of 1100 ms between the items. They were delivered at a level of approximately 70 dB.

Procedure

The selection of newborns, awakening and installation, HAS measurement, satiation criterion and experimental duration are similar to a standard HAS procedure (see, for example, Floccia, Christophe & Bertoncini, 1997). Briefly, infants were tested in a sound-attenuated chamber about 2 h after feeding. After being placed in a reclining position in a special 'baby bath chair', they were given a pacifier held by a mechanical arm to prevent any experimenter intervention.

The criterion for HA sucks was adjusted for each infant to obtain a 2 min baseline rate of about 80% of all sucks. Pre-shift stimulus was then presented contingent on the rate of HA sucks. The criterion for switching from the pre-shift to the post-shift phase was a decrease in HAS rates of 25% or more over two consecutive minutes compared with the rate in the immediately preceding minute. Once criterion was reached, presentation shifted automatically to a new stimulus for experimental subjects, and remained unchanged for the control subjects. The stimulus shift could not occur until five post-baseline minutes had elapsed, and not after 15 post-baseline minutes. The post-shift phase lasted 4 min.

Subjects

Subjects were recruited among healthy, full-term newborns at the Baudelocque Maternity Hospital in Paris, France. They suffered no complications during pregnancy or delivery, and were classified as normal after neurological evaluations on their first day of life. All subjects weighed more than 2700 g at birth, and had 5 min Apgar scores of 10.

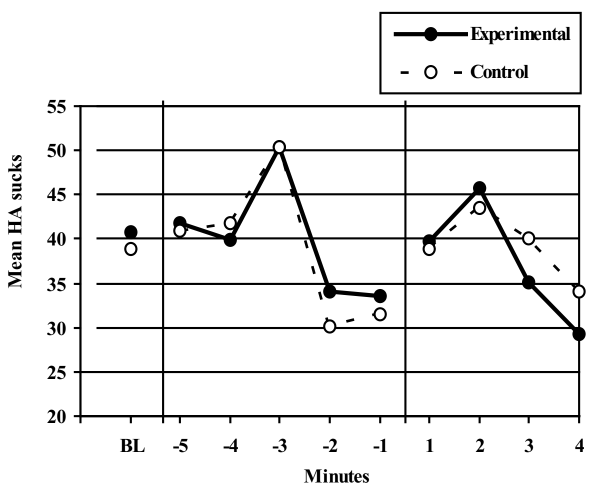
Seventy-five newborns were tested. The data from 35 newborns were excluded for the following reasons: falling asleep (14), crying or becoming fussy (nine), no HAS rate decrease criterion within 15 min (11) and experimental error (one). The remaining 40 newborns (14 females and 26 males) completed the experimental session. Their mean age was 18 h (from 8 to 24 h), mean gestational age was 39 weeks (from 37 to 41 weeks) and their mean birth weight was 3320 g (from 2700 to 4300 g). Each subject was randomly assigned to one of two conditions: experimental (20 subjects) and control (20 subjects).

Results

The plan of analyses included one between-subject factor: condition (experimental vs control).[2] Time, as a repeated measure (minute by minute), was a within-subject factor. Measure of HAS rates per minute was the dependent variable. Discrimination was evaluated by comparing HAS rates the last 2 min before the stimulus shift and the first 2 min after. The average sucking rates of both experimental and control groups are displayed in Figure 2.

Baseline and pre-shift analyses

Time for reaching the HA sucks decrease criterion was similar for the experimental and the control condition (9.2 and 9.0 min respectively; $F(1, 38) < 1$). Moreover, there was no effect of condition on HA sucks on baseline ($F(1, 38) < 1$), nor on the last 5 min of the pre-shift period ($F(1, 38) < 1$). Thus the experimental and control groups produced similar sucking behaviour before the stimulus shift.



	BL	-5	-4	-3	-2	-1	1	2	3	4
Experimental	40.8	41.8	39.8	50.4	34.0	33.5	39.6	45.7	35.1	29.2
Control	38.8	40.8	41.7	50.3	30.1	31.5	38.9	43.4	40.1	34.0

Figure 2 *Experiment 1: mean HAS rates minute by minute during baseline (BL), the last 5 min before satiation criterion (−5 to −1), and the 4 min of test (1–4), for the experimental group (straight line) and the control group (broken line).*

[2] The effect of another between-subject factor, voice (/nuga/ uttered by the female speaker as the first stimulus vs /nuga/ uttered by the male speaker), was also analysed; however, because it was found to have no significant effect in any of the analysis described below, nor did it interact with another factor, it was not included in the final plan of analyses.

Effect of stimulus change on HA sucks

To evaluate the effect of the change of stimuli, the last two pre-shift minutes were compared with the first two post-shift minutes. These comparisons revealed a significant increase in sucking rates in the experimental group (+8.9 HA sucks, $F(1, 19) = 21.04$, $p < 0.01$) as well as the control group (+10.3 HA sucks, $F(1, 19) = 13.75$, $p < 0.01$). There was no significant interaction between shift (two last pre-shift minutes vs two first post-shift minutes) and condition ($F(1, 38) < 1$), indicating that experimental and control groups behaved similarly.

Discussion of Experiment 1

In this first experiment no significant difference was observed between the experimental and control groups. In such a situation one can only hypothesize that newborns did not discriminate the voice change. Indeed the statistical and experimental power of the HAS procedure relies on the observation of a reliable difference between experimental and control groups in the after-shift period that can only be interpreted in terms of discrimination evidence. As, in the current study, a lack of difference in sucking rates does not necessarily imply that no discrimination took place, either newborns did not discriminate this voice change or the procedure did not allow a discrimination reaction to emerge. In the following section we will examine both of these possibilities, and attempt to distinguish which might explain our experimental findings.

Stimulus poverty prevented discrimination

As a first hypothesis we could suppose that newborns did not discriminate the voice change. The fact that the stimuli were short and reduced to only one token of the same word means that there could be insufficient information for discrimination of voice identity because they describe only a very small space of the variation range of each voice. Additionally, their phonetic identity results in a large overlapping of spectral composition in the two stimuli.

There are two kinds of information that could have been used to distinguish between the two productions, prosodic variations and vocal-tract related cues (timbre and $F0$ value). Prosodic variations have been considerably reduced compared with those found in a sentence-like stimulus. Intonation contours were broadly similar in both tokens (see Figure 1). Rhythmic cues and voice timbre were largely under-specified due to the use of only one token from each speaker of the same word, leading to a reduction of the available information. For instance, computing an informative distribution of each speaker syllable duration or intervals between each syllable is not possible. This might even be problematic for adult voice discrimination, for it has been shown that adults most often rely on resonant frequencies and rhythm to recognize voices (Van Dommelen, 1990).

The last cue to be considered is the *F*0 value, which was quite different across the speakers (Figure 1). However, young infants' relative failure to perceive differences in the pitch value in complex sounds has already been demonstrated. For instance, Bundy *et al.* (1982) reported that 4-month-old infants were unable to use *F*0 to recognize a change in a melody, although they could perform this task by relying on the spectral composition of harmonics (i.e. timbre). Another example is provided by Olsho (1984) who showed that, at 6 months of age, infants' discrimination sensitivity for pitch in pure tones is half that of adults for the lowest frequencies (below 1000 Hz). Given that the *F*0 range characterizing voices varies below 1000 Hz, it could be that young infants have much more difficulty than adults in discriminating differences within this range for complex sounds.

In summary, discrimination could have been impeded due to the fact that some cues are broadly similar from one token to the other (such as intonation contours and phonetic information), while some others are under-specified due to the poor variation range covered by single stimuli (such as rhythm and timbre).

Let us now consider another explanation for the absence of voice discrimination obtained in Experiment 1.

Non-sustained attention during the habituation phase

It is possible that the procedure used in the experiment was not sufficiently effective in maintaining attention during the familiarization phase, so impeding the observation of a discrimination response. First, the attention level may not have been properly maintained due to the minimal acoustic variation carried by the unchanging single token of the word /nuga/. Thus a low attentional level added to an incomplete or under-specified encoding of voice quality (because of the relatively little amount of available information in the stimulus) might explain the results found in the first experiment.

This effect could also have been amplified by the characteristics of the standard HAS procedure, in that its success relies on a single stimulus change. Subjects' attention has to be properly solicited at this precise moment so that a robust reaction can be observed. In the current experiment, as in most studies using the HAS paradigm, the stimulus shift was realized at a moment of low sucking activity, leading to few presentations of the stimulus and thus, presumably, to a decrease in attention. This may not have been problematic if the to-be-discriminated difference was important or easily encodable, but since the difference in the present study was probably less salient for newborns, then subjects were not placed in the best situation for discrimination to be observed.

In order to explore whether the absence of discrimination was due to a failure in maintaining attention throughout the habituation phase, a second experiment was carried out in which methodological changes were made. First, in order to provide more information about voice quality, multiple tokens of the word /nuga/ were used in each category (male or female). It was thought that increasing the amount of tokens might increase the space of within-voice variation, leading to more information about each voice's characteristics, whilst the context of short duration stimuli remains.

Second, in order to focus infants' attention on the voice change, multiple voice changes were introduced during the experiment. More precisely, a voice change would occur every minute, so that subjects received an alternation of female and male voice whatever their sucking activity, high or low. It was thought that by presenting the male/female voice contrast several times, light would be shed on their difference. In that respect this design is closer to preference procedures than HAS experiments, in which the to-be-discriminated stimuli are frequently alternated depending on the infant's behaviour, thus providing multiple opportunities to compare the two stimuli.

This multiple changes procedure also has statistical advantages, since it allows a reduction of the statistical importance of the traditional single change moment in the HAS procedure. Rather than one punctual data point per subject, which could be highly sensitive to attentional fluctuations, we will focus on a global effect spread over several minutes.

Experiment 2

Stimuli were similar to those used in Experiment 1, i.e. a single word uttered by two different speakers. However, in this experiment several tokens were produced by each speaker. These were alternated every minute in a Voice Change condition, compared with a No-change condition in which the stimuli remained the same throughout the session. This design is inspired from a study by

Cowan, Suomi and Morse (1982) with 2-month-old infants. If newborns discriminate the voice changes we expect HAS rates to diverge after a period according to the condition. That is, they are expected to become significantly higher in the Voice Change group than in the No-change group throughout the session (Cowan *et al.*, 1982; Floccia *et al.*, 1997; Sansavini, Bertoncini & Giovanelli, 1997).

Stimuli

The two native speakers (the same man and woman as in Experiment 1) produced six tokens of the French word /nuga/, in the same carrying sentence 'C'est /nuga/, ça?' (Is it /nuga/, this one?) so as to equalize the intonative contours as much as possible (a pause was inserted before and after the word in order to reduce coarticulation with the surrounding words). The six tokens of each category were randomly delivered during the presentation of a given category. The mean fundamental frequency of the female tokens was between 200 and 350 Hz, and between 100 and 200 Hz for the male (see Figure 3). The mean duration of the stimuli was quite similar from one talker to the other (male tokens, from 541 to 572 ms with a mean of 558 ms; female tokens, from 492 to 563 ms with a mean of 522 ms). The stimuli were sampled at 22 kHz and presented each time HA sucks were detected, with a minimal SOA of 1150 ms at a level of about 70 dB.

Design

For each condition (Voice Change and No-change), half of the subjects received the six randomly presented

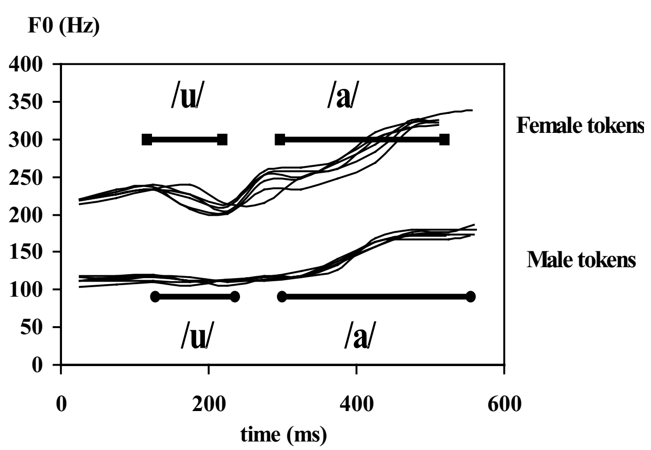


Figure 3 *Pitch contour of the female and male tokens of the word /nuga/ presented in Experiment 2. Mean duration of vowels /u/ and /a/ are reported for each voice category.*

female tokens of /nuga/ in the first minute after baseline, while the other half received the six male tokens of /nuga/ (also in a random order). In the second minute of stimulation, subjects of the No-change group continued receiving the same randomly presented six tokens, while subjects in the Voice Change group received the six tokens from the other category. This alternation was maintained for 10 min after baseline.

Procedure

The procedure of subject selection, awakening and installation, as well as the HAS measurement was similar to that used in Experiment 1. This procedure is adapted from the Cowan *et al.* study (1982) in which an experimental group of 2-month-old infants received 30 s blocks of one stimulus alternating with 30 s blocks of a second stimulus. The control group received the same stimulus throughout the session. Their results showed a consistent linear increase in sucking rates in the experimental group, but no increase in the control group, which remained at baseline level. Their interpretation was that habituation to the stimuli was delayed in the experimental condition compared with the control condition. Adapted to newborns, this alternating procedure yielded similar patterns of results: subjects presented with alternating stimuli displayed higher HAS rates than control subjects (Floccia *et al.*, 1997). However, no linear increase in HAS rates could be observed in the alternative groups, in contrast with the work with older infants. We presume that this is because newborns are unable to produce higher sucking rates than that produced during the first minutes of sucking on a blind nipple (Levin & Kaye, 1964; Floccia *et al.*, 1997). Thus, for newborns the typical pattern of discrimination observed in the alternating procedure is a decrease in HAS rates in the control groups, compared with a stable sucking activity in the alternative discriminating groups (see also Sansavini *et al.*, 1997).

Subjects

Newborns were recruited at the Port Royal Maternity Hospital in Paris and were selected with the same criteria as in the preceding experiment. However, newborns were older than in the preceding experiment: 3 days in mean compared with less than 24 h in Experiment 1.[3] Eighty-two subjects were tested. The data of 42 newborns were excluded for the following

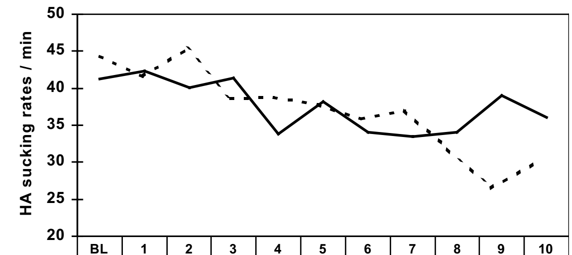
[3] Experiments 1 and 2 were carried out in two different maternity hospitals. It was not possible to test newborns younger than 2 days of life in the second maternity hospital.

reasons: falling asleep (12), ceasing to suck during the experiment (seven), crying or becoming fussy (three), rejecting the nipple (12), and having sucking rates lower than 10 HA sucks per minute during one of the first 4 min of stimulation (eight). This last criterion was introduced so that only well stimulated infants remained at the beginning of the session. The remaining 40 newborns (18 girls and 22 boys) completed the experimental session. Their mean age was 2.8 days (range 2–5 days), their mean term was 39.8 weeks (from 37.5 to 42 weeks) and their mean birth weight was 3428 g (from 2780 to 4170 g). Newborns were randomly assigned to one of the two conditions: 20 subjects in the Voice Change group and 20 in the No-change group.

Results

The plan of analyses included one between-subject factor, condition (No-change vs Voice Change), and one within-subject factor, time as a repeated measure (minute by minute).[4] The effect of condition on HA sucks was evaluated using an analysis of variance of linear trend over the 10 min of stimulation (e.g. Trehub & Chang, 1977; Cowan *et al*., 1982; Floccia *et al*., 1997). This analysis indicates whether sucking rates decrease or increase significantly over time. A decrease was expected in the No-change group, while HAS rates were expected to remain at baseline level in the Voice Change group.

The mean HAS rates minute by minute for each group are plotted in Figure 4. There was no main effect of condition on baseline ($F(1, 38) < 1$), indicating that both groups were comparable before the beginning of the stimulation period. A significant effect of condition was found in the analysis of linear trend over the 10 min of stimulation ($F(1, 38) = 2.88$, $p < 0.05$; one-tailed). As predicted, this effect was due to a significant decrease in HAS rates in the No-change group ($F(1, 19) = 11.95$, $p < 0.01$), which was more pronounced than that of the Voice Change group ($F(1, 19) = 3.24$, $p = 0.09$). This effect is mainly due to the group difference on the last 2 min of alternation ($F(1, 38) = 4.07$, $p = 0.05$). This



	BL	1	2	3	4	5	6	7	8	9	10
Voice Change	41.3	42.3	40.1	41.4	33.8	38.2	34.1	33.5	34.1	39.0	36.1
No Change	43.5	40.8	44.5	37.9	38.1	37.1	35.2	36.3	31.2	26.0	29.4

Figure 4 *HAS rates per minute for the Voice Change group (straight line) and the No-change group (broken line). BL refers to the mean of HA sucks in the 2 min baseline.*

pattern of discrimination results is similar to that obtained in previous studies with the same procedure and the same population (Floccia *et al*., 1997; Sansavini *et al*., 1997), as well as with 2-month-old infants (Cowan *et al*., 1982).

Discussion of Experiment 2

The results of Experiment 2 show that newborns stimulated by the alternation of talker-changing stimuli exhibit significantly different behaviour compared to those of the control group. These results can be taken as evidence of speaker change discrimination in a single disyllabic word context.

Contrary to the first hypothesis presented for the explanation of the results of Experiment 1, the results of this experiment suggest that isolated disyllabic words are not too short to allow newborns to process their voice quality. Thus the negative results in Experiment 1 cannot be attributed to newborns' intrinsic failure to extract a speaker-specific pattern from short speech sounds, as they display such a capacity in Experiment 2.

The main interpretation of the different outcomes of Experiments 1 and 2 is that the use of a multiple changes procedure, and an increase in the number of tokens, contributed, in different ways, to attract newborns' attention toward the processing of voice differences. Before discussing this point, two other possibilities must be considered, the first connected with the acoustic properties of the stimuli used in the two experiments, and the second related to subjects' age.

In both experiments, tokens of the word /nuga/ produced by the same male and female talkers were used. Despite their great similarity, the tokens used in the two experiments differed in one major characteristic; i.e. the intonation contour of the tokens used in Experiment 2 was rising, whilst in Experiment 1 it was falling. Could it be that a rising intonative contour

[4] Another factor was the speaker voice (the female tokens or the male tokens): it was a between-subject factor for the two No-change subgroups (half of them received the male tokens, the other half received the female tokens), and a within-subject factor for the two Voice Change subgroups (both received an alternation of female and male tokens). In none of the analyses detailed below did voice have a significant effect; nor did it interact with another factor. In particular, it was verified that the two No-change subgroups did not differ from each other on the 10 min of stimulation ($F(1, 18) < 1$), and that newborns in the two Voice Change subgroups did not suck significantly more during the presentation of the female tokens than during that of the male tokens ($F(1, 18) = 1.86$). Thus this factor was not included in the final plan of analyses.

better attracts newborns' attention than a falling one, hence allowing for a more specified representation of the to-be-discriminated items in Experiment 2? Sullivan and Horowitz (1983) investigated differential attention for falling versus rising contour in infants aged from 2 to 3 months. An infant-control auditory-preference paradigm was used to assess subjects' attention to synthetically generated and naturally produced rising and falling intonation contours in isolated syllables. Results revealed that infants showed greater attention to the naturally produced rising intonation contour. However, a reverse pattern of greater attention to the falling contour was found with the synthetically generated stimuli. In addition, rising contours in sentences seem to be used preferentially by caretakers to elicit eye contact with infants, while sinusoidal and bell-shaped contours are used to maintain the infant's positive affect and gaze (Stern, Spieker & MacKain, 1982). Thus the rising contour of Experiment 2's stimuli might be an additional factor that contributed, together with the changes we introduced, to keep attention at a level appropriate for discrimination to occur.

A second factor that could partially explain the difference in results of Experiments 1 and 2 is the age difference between the subjects: the mean age of newborns in the first experiment was less than 24 h, while it was 2.8 days in Experiment 2. This difference could have several consequences: (a) younger infants could be less effective than older infants in learning the contingency between their HA sucks and the presentation of sounds; they could also have less accurate perceptual capacities for speech sounds because of differences in (b) maturation of their auditory system, (c) exposure to linguistic stimuli, and/or (d) peripheral problems such as a higher auditory threshold. As we shall argue, we believe that it is only the last point which could have a serious effect upon the results of this experiment.

Regarding point (a), Fifer (1981) reported no difference in response to tone/no tone signals between 1- and 3-day-old newborns in an operant conditioning procedure also using non-nutritive sucking. In a similar study, using syllables as signals, an effect of age was found between newborns aged 1–2 days and older ones, in that there was a greater response of younger rather than older infants (Moon & Fifer, 1990).

Regarding maturation of the auditory system (point (b)), there is more variation in maturation within each newborn group – the 1-day-old and the 3-day-old groups – than between them, because of the variability in infants' gestational term (from 37 weeks to 41 weeks in each group).

Finally, regarding linguistic exposure (point (c)), *post hoc* analyses revealed no significant correlation between the duration of subjects' post-natal life[5] and their recovery in HA sucks in Experiment 1 ($r = -0.26$; $p = 0.26$), and no interaction between age and condition (Voice Change vs No-change) in the HA sucks linear trend in Experiment 2 ($F(1, 36) < 1$).

The most serious argument related to subjects' age is probably that of the high auditory threshold during the first hours or days after birth (point (d)). At birth the middle ear still contains fluid and some fragments can be found in the external ear canal, hindering the sound transmission to the inner ear. A few hours, or even days, may be necessary for the external and middle ears to be cleared. Thus the youngest group of newborns may have, on average, an auditory threshold higher than that of the eldest subjects (T. Morlet, personal communication, 8 April 1999). However, measures of auditory evoked potentials soon after birth reveal no significant differences across newborns of different ages, at least for the few days following delivery (e.g. Kurtzberg, Hilpert, Kreuzer & Vaughan, 1984).

Rather than subjects' age, we hypothesize that it is both the stimulus intonative contour differences and the changes in stimulus presentation, in terms of number of tokens and voice switches, that lead to different attentional demands in Experiments 1 and 2, resulting in different discrimination outcomes.

Why would an increase in the number of tokens and voice switches change the level of attention? First, both changes contributed to introduce variability in stimuli compared with Experiment 1. Increasing variability in stimulation generally leads to an arousal of attention. For instance, early preference for motherese is mainly explained by the greater variability of *F*0 compared with adult-directed speech (Fernald & Kuhl, 1987) that contributes to attracting infant's attention toward this specific input (e.g. Cooper & Aslin, 1989; Trainor, Clark, Huntley & Adams, 1997), greater attention being linked with better memorization of the items (Adler, Gerhardstein & Rovee-Collier, 1998).

However, the aim was not only to add random variability in the stimulus set: the organization of stimulus presentation was to be considered carefully. Bornstein (1981) has already pointed out that, in the visual perception domain, the results obtained in habituation and dishabituation phases in infants were dependent on the nature of the variation in stimuli presented during those periods. In the auditory perception domain, a few studies using HAS in infants or newborns have questioned the influence of stimulus

[5] Subjects' age is a poor predictor of the amount of time they have spent hearing speech sounds between birth and the experimental session. Thus this analysis of correlation must be taken cautiously.

organization upon the attentional resources involved in this procedure (Kuhl & Miller, 1982; Jusczyk, Bertoncini, Bijeljac-Babic, Kennedy & Mehler, 1990; Jusczyk, Pisoni & Mullenix, 1992). From these studies, it appears that discrimination results depend on how the stimulus variability presented is distributed during the familiarization phase. For example, in Jusczyk *et al.* (1990), infants (newborns and 2-month-olds) reacted differently to the adjunction of a novel syllable in a set of syllables, as a function of the acoustic distance between this new item and the familiar ones as well as of the distance between each item within the set of familiar syllables. This kind of finding is not surprising given that psychological space can be modified depending on stimulus organization and variability, at least in adults (e.g. Nosofsky, Clark & Shin, 1989). Methodological changes in Experiment 2 contributed to enhancing infants' discrimination performances compared with Experiment 1 because it provided them with the opportunity to categorize voices, i.e. to perceive that there is less variation amongst tokens uttered by the same voice than between tokens uttered by the two voices. The within-category similarity was emphasized by presenting multiple tokens of the same voice, while the between-category dissimilarity was emphasized by alternating the voice every minute, leading to an increase in the ease with which newborns could recognize the categorical voice distinction (see Madole & Oakes, 1999).

Concluding remarks

This study clearly demonstrates the need for further research on the systematic investigation of the relations between experimental procedures and underlying mechanisms of perceptual processing in infants in the auditory domain. In particular it will be necessary to examine how the level of attention varies as a function of the organization of stimuli presented to the subjects, and the consequences of this upon discrimination or preference responses.

Finally, we have shown that newborns are able to discriminate unfamiliar voices given a minimal (short duration) information context. Future research should focus on the role these items of acoustic information play in the construction of infants' lexical representations. Most theories of spoken word identification assume that variable speech signals are matched to canonical representations in memory. To achieve this, idiosyncratic details such as voices are normalized before allowing direct comparison of the input to the lexicon. For instance, event-related potentials recorded during recognition tasks for spoken words alone or for both words and the voice of the speaker indicate that word and voice information are retrieved hierarchically in adults (Senkfor & Van Petten, 1998). In addition, voice recognition capacities can be specifically impaired as is found in phonagnosia (Van Lancker, Kreiman & Cummings, 1989). However, recent evidence suggests that voice recognition and the phonetic content of a linguistic utterance are not independent. Memory tests for spoken words as a function of speakers' voices suggest that episodic memory traces of spoken words retain the surface details typically considered as noise in perceptual systems (Goldinger, 1996; see also Nygaard, Sommers & Pisoni, 1994; Sheffert, 1998). Showing that newborns can discriminate speaker change in an isolated-word context is a first step in the determination of how this speaker-related information is encoded and used in early lexical acquisition.

Acknowledgements

This research was performed when all three authors belonged to the Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS-EHESS, Paris, under the supervision of Jacques Mehler. The first experiment is part of Sylvaine Rousseau's DEA (Université Paris VI). Many thanks to Jeremy Goslin and Thierry Morlet.

References

Abberton, E., & Fourcin, A.J. (1978). Intonation and speaker identification. *Language and Speech*, **21** (4), 305–318.

Adler, S.A., Gerhardstein, P., & Rovee-Collier, C. (1998). Levels of processing effects in infant memory. *Child Development*, **69** (2), 280–294.

Bertoncini, J., Morais, J., Bijeljac-Babic, R., McAdams, S., Peretz, I., & Mehler, J. (1989). Dichotic perception and laterality in neonates. *Brain and Language*, **37**, 591–605.

Best, C.T., Hoffman, H., & Glanville, B.B. (1982). Development of infant ear asymmetries for speech and music. *Perception and Psychophysics*, **31**, 75–85.

Bornstein, M.H. (1981). Psychological studies of color perception in human infants: habituation, discrimination and categorization, recognition, and conceptualization. *Advances in Infancy Research*, **1**, 1–40.

Bundy, R.S., Colombo, J., & Singer, J. (1982). Pitch perception in young infants. *Developmental Psychology*, **18**, 10–14.

Clarkson, M.G., Clifton, R.K., & Perris, E.E. (1988). Infant timbre perception: discrimination of spectral envelopes. *Perception and Psychophysics*, **43**, 15–20.

Cooper, R.P., & Aslin, R.N. (1989). The language environment of the young infant: implications for early perceptual development. *Canadian Journal of Psychology*, **43** (2), 247–265.

Cowan, N., Suomi, K., & Morse, P.A. (1982). Echoic storage in infant perception. *Child Development*, **53**, 984–990.

Crowder, R.G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds), *Thinking in Sounds: The Cognitive Psychology of Audition*. Oxford: Oxford University Press.

DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: newborns prefer their mother's voices. *Science*, **208**, 1174–1176.

DeCasper, A.J., & Prescott, P.A. (1984). Human newborns' perception of male voices: preference, discrimination, and reinforcing value. *Developmental Psychobiology*, **17**, 481–491.

Endres, W., Bambach, W., & Flösser, G. (1971). Voice spectrograms as a function of age, voice disguise and voice imitation. *Journal of the Acoustical Society of America*, **49**, 1842–1848.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, **8**, 181–195.

Fernald, A., & Kuhl, P.K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, **10** (3), 279–293.

Fifer, W.P. (1981). Early attachment: maternal voice preference in one- and three-day-old infants. Unpublished doctoral dissertation, University of North Carolina, Greensboro.

Floccia, C., Christophe, A., & Bertoncini, J. (1997). High amplitude sucking and newborns: the quest for underlying mechanisms. *Journal of Experimental Child Psychology*, **64**, 175–198.

Goggin, J.P., Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of familiarity in voice identification. *Memory and Cognition*, **19**, 448–458.

Goldinger, S.D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22** (5), 1166–1183.

Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P.W., Bertoncini, J., Bijeljac-Babic, R., Kennedy, L.J., & Mehler, J. (1990). The role of attention in speech perception by young infants. *Cognitive Development*, **5**, 265–286.

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.

Karzon, R.G., & Nicholas, J.G. (1989). Syllabic pitch perception in 2- to 3-month-old infants. *Perception and Psychophysics*, **45**, 10–14.

Kuhl, P.K., & Miller, J.D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics*, **31**, 279–292.

Kurtzberg, D., Hilpert, P.L., Kreuzer, J.A., & Vaughan, H.G., Jr (1984). Differential maturation of cortical auditory evoked potentials to speech sounds in normal fullterm and very low-birthweight infants. *Developmental Medicine and Child Neurology*, **26**, 466–475.

Lecanuet, J.P., Granier-Deferre, C., Jacquet, A.-Y., Capponi, I., & Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. *Early Development and Parenting*, **2**, 217–228.

Levin, G.R., & Kaye, H. (1964). Nonnutritive sucking by human neonates. *Child Development*, **35**, 749–758.

Madole, K.L., & Oakes, L.M. (1999). Making sense of infant categorization: stable processes and changing representations. *Developmental Review*, **19**, 263–296.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, **7**, 491–497.

Moon, C., & Fifer, W.P. (1990). Syllables as signals for 2-day-old infants. *Infant Behavior and Development*, **13**, 377–390.

Morgan, J.L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, **35** (5), 666–688.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24** (3), 756–766.

Nosofsky, R.M., Clark, S.E., & Shin, H.J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 282–304.

Nygaard, L.C., Sommers, M.S., & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5** (1), 42–46.

Olsho, L.W. (1984). Infant frequency discrimination. *Infant Behavior and Development*, **7**, 27–35.

Remez, R.E., Fellowes, J.M., & Rubin, P.E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, **23** (3), 651–666.

Sansavini, A., Bertoncini, J., & Giovanelli, J. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, **33**, 3–11.

Senkfor, A.J., & Van Petten, C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24** (4), 1005–1025.

Sheffert, S. (1998). Contributions of surface and conceptual information to recognition memory. *Perception and Psychophysics*, **60** (7), 1141–1152.

Shen, Z., Lin, S., & Wu, P. (1992). Monkey ERPs during identification of familiar voices. *Acta Psychologica Sinica*, **24** (4), 400–405.

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**, 199–212.

Stern, D.N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, **18** (5), 727–735.

Sullivan, J.W., & Horowitz, F.D. (1983). The effects of intonation on infant attention: the role of the rising intonation contour. *Journal of Child Language*, **10** (3), 521–534.

Trainor, L.J., Clark, E.D., Huntley, A., & Adams, B.A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development*, **20** (3), 383–396.

Trehub, S.E., & Chang, H.W. (1977). Speech as a reinforcing stimulation for infants. *Developmental Psychology*, **13**, 170–171.

Trehub, S.E., Endman, M.W., & Thorpe, L.A. (1990). Infants' perception of timbre: classification of complex tones by spectral structure. *Journal of Experimental Child Psychology*, **49**, 300–313.

Trehub, S.E., Thorpe, L.A., & Morrongiello, B.A. (1985). Infants' perception of melodies: changes in a single tone. *Infant Behavior and Development*, **8**, 213–223.

Van Dommelen, W.A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, **33**, 259–272.

Van Lancker, D.R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology*, **11** (5), 665–674.

Wormith, S.J., Pankhurst, D., & Moffitt, A.R. (1975). Frequency discrimination by young infants. *Child Development*, **46**, 272–275.

Received: 7 June 1999
Accepted: 11 February 2000