



HAL
open science

Perception of Speech Modulation Cues by 6-Month-Old Infants

Laurianne Cabrera, Josiane Bertoncini, Christian Lorenzi

► **To cite this version:**

Laurianne Cabrera, Josiane Bertoncini, Christian Lorenzi. Perception of Speech Modulation Cues by 6-Month-Old Infants. *Journal of Speech, Language, and Hearing Research*, 2013, 56 (6), pp.1733. hal-01968824

HAL Id: hal-01968824

<https://hal.science/hal-01968824v1>

Submitted on 3 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perception of speech modulation cues by 6-month-old infants

Cabrera Laurianne^{1,2}, Bertoncini Josiane¹, and Lorenzi Christian^{1,2}

¹ *Laboratoire de Psychologie de la Perception*

CNRS, Université Paris Descartes

45 rue des saints Pères, 75006 Paris, France

² *Département d'Etudes Cognitives, Institut d'Etude de la Cognition*

Ecole normale supérieure, Paris Sciences et Lettres

29 rue d'Ulm, 75005 Paris, France

Abstract

Purpose: This study assessed the capacity of 6-month-old infants to discriminate a voicing contrast (/aba/-/apa/) on the basis of amplitude modulation cues (AM, the variations in amplitude over time within each frequency band) and frequency modulation cues (FM, the oscillations in instantaneous frequency close to the center frequency of the band).

Method: Several vocoded speech conditions were designed to: (i) degrade FM cues in 4 or 32 bands, or (ii) degrade AM in 32 bands. Infants were familiarized to the vocoded stimuli for a period of either 1 or 2 min. Vocoded speech discrimination was assessed using the head-turn preference procedure.

Results: Infants discriminated /aba/ from /apa/ in each condition. However, familiarization time was found to influence strongly infants' responses (*i.e.*, their preference for novel *versus* familiar stimuli).

Conclusions: Six-month-old infants do not require FM cues, and can use the slowest (<16 Hz) AM cues to discriminate voicing. Moreover, six-month-old infants can use AM cues extracted from only four broad frequency bands to discriminate voicing.

Key Words: Speech perception, Vocoder, Modulation cues, Normal-hearing infants

Introduction

Most recent studies about speech perception in infants have focused on how they acquire the phonological properties of their native language in a variety of learning contexts (see Kuhl *et al.*, 2008, for a review). The present study takes a different approach by exploring how infants process the acoustic information related to phonetic differences. More precisely, this study aims to assess to which extent infants rely on low-level (*i.e.*, sensory) spectro-temporal *modulation* cues to discriminate a voicing contrast (see also Bertoncini, Nazzi, Cabrera, & Lorenzi, 2011).

The auditory mechanisms underlying speech perception in adulthood have been thoroughly described. Speech signals are decomposed by the auditory filters in the cochlea into many narrow frequency bands with a passband equal to one “equivalent-rectangular bandwidth”, 1-ERB_N (Glasberg & Moore, 1990; Moore, 2003). Over the last two decades, several psychophysical studies reconsidered speech-perception processes based on the general assumption that each 1-ERB_N wide band of speech should be viewed as a sinusoidal carrier with superimposed amplitude modulation (AM) and frequency modulation (FM) (*e.g.*, Drullman, 1995; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Sheft, Ardoint, & Lorenzi, 2008; Smith, Delgutte, & Oxenham, 2002; Zeng *et al.*, 2005). The AM component – often referred to as the “acoustic temporal envelope” - corresponds to the relatively slow modulations in amplitude over time. The FM component – often referred to as the “acoustic temporal fine structure” - represents the relatively fast fluctuations in instantaneous frequency over time with average frequency close to the center frequency of the 1-ERB_N wide band.

Both AM and FM features are represented in the phase-locking pattern of auditory-nerve fibers’ discharges. For most adult mammals, the accuracy of neural phase locking to instantaneous frequency (and thus, FM cues) is constant up to about 1-2 kHz and then

1
2
3
4
5 declines so that phase locking is no longer detectable at about 5-6 kHz (*e.g.*, Johnson, 1980;
6
7 Kiang, Pfeiffer, & Starr, 1965; Palmer & Russell, 1986; Rose, Brugge, Anderson, & Hind,
8
9 1967). In contrast, neural phase locking to AM cues remains accurate for carrier (audio)
10
11 frequencies well beyond 6 kHz (*e.g.*, Joris, Schreiner, & Rees, 2004; Joris & Yin, 1992; Kale
12
13 & Heinz, 2010). Moreover, several additional peripheral auditory mechanisms such as
14
15 synaptic adaptation at the inner hair-cell level appear to limit AM coding beyond those that
16
17 limit instantaneous-frequency coding (Joris & Yin, 1992). These physiological data suggest
18
19 some form of dissociation in the representation of AM and FM features in the early stages of
20
21 the auditory system. However, it has been proposed that only the slowest FM (<5-10 Hz) cues
22
23 are encoded in a purely temporal manner (that is, independently of AM) via phase locking in
24
25 auditory-nerve fibers whereas faster FM cues (>10 Hz) are encoded via a place (*i.e.*,
26
27 tonotopic) mechanism (*e.g.*, Moore & Sek, 1996; Saberi & Hafter, 1995)¹.
28
29
30
31

32 Phonetic features such as voicing, nasality, place and manner of articulation are signaled
33
34 by various AM cues (*e.g.*, the existence and duration of silent intervals are important in
35
36 distinguishing voiced from voiceless plosives in intervocalic position; Rosen, 1992), which
37
38 are relatively widespread or located in the high audio-frequency range (Rosen, 1992).
39
40 Segmental cues to voicing and nasality are restricted to the low-mid audio frequency range
41
42 (*e.g.*, voiced sounds have a power spectrum heavily weighted to audio frequencies below 1
43
44 kHz; Rosen, 1992) and are well represented in the pattern of phase locking in auditory-nerve
45
46 fibers (*e.g.*, Deng & Geisler, 1987; Sinex & Geisler, 1983), suggesting that voicing and
47
48 nasality may also be signaled by FM cues encoded in a purely temporal manner.
49
50
51
52
53
54

55
56 ¹ In this case, the differential attenuation of cochlear filtering converts the frequency excursions of FM into AM
57
58 fluctuations at the output of auditory filters in the cochlea, a process sometimes referred to as “temporal
59
60 envelope reconstruction” (see Gilbert & Lorenzi, 2006; Zeng *et al.*, 2004, for applications to speech perception).

1
2
3
4
5 A number of speech perception studies have attempted to assess the relative importance
6
7 of AM and FM cues in speech identification and phonetic-feature perception for normal-
8
9 hearing adults. These studies have used nonsense syllables, words or sentences processed by
10
11 vocoders, that is signal-processing algorithms which extract and alter selectively AM and FM
12
13 cues within specific analysis frequency bands (Dudley, 1939). After filtering out or
14
15 scrambling the original AM or FM speech cues within each analysis frequency band, the
16
17 resulting signal is assumed to retain mainly the AM or the FM speech cues (*e.g.*, Gilbert &
18
19 Lorenzi, 2006; Shannon *et al.*, 1995; Sheft *et al.*, 2008; Smith *et al.*, 2002; Zeng *et al.*, 2005;).
20
21 These studies indicated that normal-hearing adults can achieve high levels of speech
22
23 intelligibility and phonetic-feature perception with speech stimuli vocoded to retain mainly
24
25 the AM speech cues (stimuli referred to as “AM speech” thereafter). This was initially
26
27 demonstrated by Shannon *et al.* (1995) who evaluated English-speaking adults’ abilities to
28
29 identify nonsense syllables in different speech-processing conditions. The AM-speech stimuli
30
31 contained only AM cues extracted within a limited number of broad analysis frequency bands
32
33 (1, 2, 3 or 4 bands) using a lowpass filter with a cutoff frequency varying from 16 to 500 Hz.
34
35 Syllable identification and phonetic-feature perception (*i.e.*, voicing, manner and place of
36
37 articulation) were poor for 1, 2, or 3 analysis frequency bands, but sharply increased with 4
38
39 bands, irrespective of the cutoff frequency of the lowpass filter used to extract AM. However,
40
41 in order to reach such high levels of accuracy, participants required a relatively long training
42
43 period (8 to 10 hours).
44
45
46
47
48

49 Other studies with vocoded signals revealed that adult listeners could also achieve high
50
51 levels of speech perception with nonsense syllables or sentences while the physical signal
52
53 retains mainly the FM cues (*e.g.*, Gilbert & Lorenzi, 2006; Hopkins, Moore, & Stone, 2010;
54
55 Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Sheft *et al.*, 2008). However, participants
56
57
58
59
60

1
2
3
4
5 required a much longer training period to identify accurately these “FM-speech” stimuli than
6
7 their “AM-speech” counterparts. These data suggested that AM cues play a more important
8
9 role than FM cues in accurate speech recognition. However, a number of studies conducted
10
11 with adults tested with vocoded speech demonstrated that the relative importance of FM cues
12
13 may increase when AM cues are degraded by various acoustic distortions (*e.g.*, Ardoint &
14
15 Lorenzi, 2010; Gilbert, Bergeras, Voillery, & Lorenzi, 2007; ; Gilbert & Lorenzi, 2006;
16
17 Hopkins, Moore, & Stone, 2008; Nelson, Jin, Carney, & Nelson, 2003; Qin & Oxenham,
18
19 2003; Zeng *et al.*, 2005). Indeed, speech perception was found to be poorer for “AM speech”
20
21 than for “intact speech” (that is, for speech combining AM and FM cues within each
22
23 frequency band) when stimuli were spectrally reduced or filtered, periodically interrupted or
24
25 masked by interfering talkers or background noise (Eaves, Summerfield, & Kitterick, 2011;
26
27 Gnansia, Pean, Meyer, & Lorenzi, 2009).

28
29
30
31
32 These data demonstrate that for adults, both AM and FM cues convey phonetic
33
34 information whose relative weight may vary according to the listening conditions, and in a
35
36 more general sense, as a function of speech redundancy. Thus, for adults, the loss of AM cues
37
38 may be compensated for by relying more on FM cues. While numerous studies have
39
40 investigated speech perception in adults, information is still lacking regarding the ability of
41
42 neonates, infants and older children to use AM and FM cues in speech. Do infants use AM
43
44 and FM cues in the same way as adults when listening to speech sounds? Do the acoustic
45
46 degradations of AM and FM cues impair the performance of developing auditory mechanisms
47
48 as much as that of mature (coupled) auditory and speech mechanisms? To our knowledge,
49
50 only a few studies have directly addressed the developmental course of modulation perception
51
52 using vocoded speech stimuli.
53
54
55
56
57
58
59
60

1
2
3
4
5 In a pioneering study, Eisenberg, Shannon, Shaefer Martinez, Wygonski and Boothroyd
6
7 (2000) assessed the ability of normal-hearing 7- and 10-year-old children and adults to identify
8 nonsense syllables, words and sentences vocoded to retain only AM cues below 160 Hz
9 within 4 to 8 analysis frequency bands (the FM carriers were replaced by noise in each band).
10
11 The results showed that children less than 7 years required a higher frequency resolution (*i.e.*,
12 a greater number of analysis frequency bands) than 10 year-olds and adults to reach similar
13 identification performance with the AM-speech stimuli. Interestingly, voicing perception (and
14 especially, perception of the /sa/ versus /za/ contrast) was the poorest across all subjects
15 groups.
16
17

18
19
20
21
22
23
24
25 This initial investigation was extended by Bertoncini, Serniclaes and Lorenzi (2009) to
26 younger children aged between 5 and 7 years. A discrimination task was used with nonsense
27 syllables vocoded to retain only AM cues below 64 Hz within 16 frequency bands (here, the
28 FM carriers were replaced by pure tones with fixed frequencies). Bertoncini *et al.* (2009)
29 found that normal-hearing 5-, 6-, and 7-year-old children were able to discriminate speech
30 contrasts on the basis of AM cues at an adult level. This study also showed no significant
31 difference in performance across age groups for voicing, place, manner, and nasality. Thus,
32 the perception of AM cues below 64 Hz in speech appeared as robust for 5-year-old children
33 as for adults when a discrimination task was used. However, consistent with the results of
34 Eisenberg *et al.* (2000), children's discrimination scores were lowest for the voicing contrast.
35 This finding was attributed to the attenuation of fundamental-frequency (F0) energy at the
36 onset of voicing caused by the removal of FM cues.
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51
52 Bertoncini *et al.* (2011) further explored the perception of voicing based on AM cues by
53 testing six-month-old infants. A behavioral task (based on the head-turn preference
54 procedure) was used to assess the ability of infants to discriminate pairs of AM-vocoded
55
56
57
58
59
60

1
2
3
4
5 syllables such as /aba/ and /apa/. As in Bertoncini *et al.* (2009), the AM cues were extracted
6
7 within 16 analysis frequency bands and lowpass filtered at 64 Hz. Again, the original FM
8
9 carriers were replaced by pure tones. The results showed that infants attended significantly
10
11 longer to alternating sequences of /aba/ and /apa/ than to repeated sequences of either /aba/ or
12
13 /apa/. This suggested that infants detected the alternation on the basis of the AM cues
14
15 differentiating voicing information.
16
17

18
19 These studies revealed that infants and older children are able to discriminate phonetic
20
21 features on the sole basis of AM cues. However, Eisenberg *et al.* (2000)'s study suggested
22
23 that the capacity to resist degradations of AM and/or FM cues in speech sounds in a more
24
25 demanding identification task may not be entirely mature before at least the age of 7 years.
26
27 This is compatible with the outcome of several psychophysical studies conducted with non-
28
29 linguistic stimuli, showing that auditory sensitivity to AM and FM cues is not adult-like until
30
31 around 10-11 years of age (*e.g.*, Aslin, 1989; Colombo & Horowitz, 1986; Hall & Grose,
32
33 1994; Moore, Cowan, Riley, Edmondson-Jones, & Ferguson, 2011; for a review, see also
34
35 Saffran, Werker, & Werner, 2006, and Werner & Gray, 1998). Nonetheless, other studies
36
37 conducted both in infants (*e.g.*, Abdala & Folsom, 1995; Levi, Folsom, & Dobie, 1995;
38
39 Olsho, 1985; Spetner & Olsho, 1990) and in cats (Brugge, Javel, & Kitzes, 1978; Kettner,
40
41 Feng, & Brugge, 1985) suggest that some aspects of frequency selectivity and neural phase
42
43 locking in auditory-nerve fibers and brain-stem neurons should be mature from early in
44
45 infancy. Regarding temporal resolution (that is the ability to follow changes in AM
46
47 fluctuations as a function of time), the results vary according to the nature of the stimuli and
48
49 according to the methods. In some studies, temporal resolution approaches an adult-like
50
51 profile at about 6 months of age (*e.g.*, Levi & Werner, 1996; Trainor, Samuel, Desjardins, &
52
53 Sonnadara, 2001; Trehub, Schneider, & Henderson, 1995). At the same time, other studies
54
55
56
57
58
59
60

1
2
3
4
5 show differences across age groups, indicating that temporal sensitivity is generally poorer in
6
7 infants and more dependent on the task or on sound complexity than in adults (Buss, Hall,
8
9 Grose, & Dev, 1999; Diedler, Pietz, Bast, & Rupp, 2007; Smith, Trainor, & Shore, 2006;
10
11 Wightman, Allen, Dolan, Kistler, & Jamieson, 1989). Nevertheless, it is generally admitted
12
13 that, under some circumstances, the mechanisms governing auditory temporal resolution in
14
15 infants operate qualitatively like those in adults (Werner, Marean, Halpin, Spetner, &
16
17 Gillenwater, 1992).
18
19

20
21 The above review of infants' abilities to process spectral and temporal auditory cues
22
23 reveals important disparities between psychophysical studies using non-linguistic stimuli
24
25 (*e.g.*, pure tones, complex tones, noise bursts) and speech-acquisition studies using linguistic
26
27 stimuli (*e.g.*, syllables, words). The present study attempted to address these disparities by
28
29 combining psychophysical and psycholinguistic methods to explore 6-month-old infants'
30
31 abilities to process phonetic information on the basis of AM and FM speech cues. More
32
33 precisely, the present study investigated whether or not 6-month-old infants discriminate a
34
35 French voicing contrast (*/aba/ versus /apa/*) in spite of degradations of the AM and/or FM
36
37 cues in speech sounds.
38
39

40
41 An impressive number of studies on the development of speech perception have shown
42
43 that, during the first months of life, infants can discriminate many different phonetic contrasts,
44
45 including non-native ones (*e.g.*, Mattock, Molnar, Polka, & Burnham, 2008; for a review see
46
47 Kuhl, 2004, and Werker & Tees, 1999). Voicing discrimination, and more precisely
48
49 discrimination of the */b/ versus /p/* contrast, was demonstrated in infants as young as one
50
51 month of age by Eimas, Siqueland, Jusczyk and Vigorito (1971) using the high-amplitude
52
53 sucking method. However, at the end of the first year, speech perception is marked not only
54
55 by a decline in non-native contrast discrimination (*e.g.*, Werker & Tees, 1983) but also by
56
57
58
59
60

1
2
3
4
5 facilitation in native contrast discrimination (*e.g.*, Kuhl *et al.*, 2006). Moreover, a recent study
6
7 indicated that between 4- and 8-months of age, French-learning infants become more sensitive
8
9 to the French value (0 ms) of the voice-onset-time (VOT) boundary between voiced and
10
11 voiceless plosive consonants (Hoonhorst *et al.*, 2009). Therefore, the speech-processing
12
13 mechanisms responsible for the perception of phonetic features such as voicing are developed
14
15 at 6 months of age, but far from being entirely tuned to the typical contrastive patterns of the
16
17 native language (*e.g.*, Mattock *et al.*, 2008; Werker & Tees, 1983, but see Kuhl, Williams,
18
19 Lacerda, Stevens, & Lindblom, 1992).
20
21

22
23 The present study explored the early capacity of 6-month-old French-learning infants to
24
25 use AM and FM cues in discriminating a voicing contrast. Two sets of syllables (/aba/ and
26
27 /apa/ stimuli) were either left intact or processed by a multi-channel, noise-excited vocoder in
28
29 order to: (i) degrade FM cues by replacing the FM carriers by bands of noise in each analysis
30
31 frequency band, (ii) degrade AM cues – and especially F0-related periodic AM cues - by
32
33 filtering out the AM components above 16 Hz in each frequency band, and (iii) degrade AM
34
35 cues by reducing the frequency resolution of the vocoder (*i.e.*, the number of analysis
36
37 frequency bands) from 32, 1-ERB_N wide to 4, 8-ERB_N wide frequency bands. These three
38
39 vocoded conditions were designed to assess whether normal-hearing 6-month-olds are able to
40
41 discriminate a voicing contrast (/aba/ *versus* /apa/) on the sole basis of: (i) AM speech cues,
42
43 (ii) the slowest (< 16 Hz) AM speech cues, and (iii) the AM cues extracted from a limited
44
45 number of broad frequency bands.
46
47
48

49
50 The intact and processed speech stimuli were presented to 6-month-old infants and an
51
52 adaptation of the head turn preference procedure (HPP) was used to assess their
53
54 discrimination ability. The procedure used in this experiment was adapted from the one
55
56 introduced by Hirsh-Pasek *et al.* (1987). Our modified version of the HPP included a
57
58
59
60

familiarization phase during which a stimulus is presented, followed by a test phase in which sequences of the (same) familiar stimulus alternated with sequences of a (different) novel stimulus. This version of HPP has been used recently in a study on melody discrimination in 2-month-old infants (Plantinga & Trainor, 2009) and in several psycholinguistic studies assessing speech discrimination for 5-to-9-month-olds (*e.g.* Bosch & Sebastian-Gallés, 2001; Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Nazzi, Juszczyk, & Johnson, 2000; Skoruppa *et al.*, 2009). These studies showed that the familiarization period could induce in infants a preference for the familiar or novel sequences *if and only if* infants are able to discriminate these sequences². In most studies, the duration of familiarization was set to 1 min but in some cases this duration was extended to 2 min (*e.g.* Bijeljac-Babic, Serres, Höhle, & Nazzi, 2012). In the present experiment, the duration of the familiarization phase was first set to 1 min. For each experimental condition, we expected that the novel sequences would yield longer (correctly oriented) looking times than the familiar sequences.

Experiment 1:

Participants

Six-month-old infants were recruited from a database of birth announcements. All families were informed about the goals of the current study and provided a written consent before their participation, in accordance with current French ethical requirements.

Data from 88 infants from French-speaking families (22 infants per condition) were analysed in this experiment (40 boys and 48 girls; age range: 5 months 26 days; 7 months 10

² The procedure differs from a “spontaneous preference” paradigm in that the familiarization was achieved with sound A for half of the participants and with sound B for the other half, in such a way that preference for novel or for familiar stimuli cannot be confused with preference for one particular category (A or B). Second, it should be noted that while discrimination responses do not imply preference, displaying a preference necessitates discrimination between the two types of sequences.

1
2
3
4
5 days; mean = 6 months 16 days; $SD = 9$ days). All infants are normal-hearing (based on
6
7 parental report of newborn hearing screening results). The data from 37 additional infants
8
9 were not included for the following reasons: fussing and crying ($n=22$), looking time shorter
10
11 than 1500 ms for one trial ($n=6$), outlier number of familiarization trials ($n=6$) and extreme
12
13 mean looking times leading to outlier differences (more than, or less than the mean difference
14
15 plus or minus 2 SD , respectively) between novel and familiar series ($n=3$).
16
17

18 Stimuli

19
20 Speech signals were recorded in a soundproof room and digitized (16-bit resolution) at a
21
22 44.1-kHz sampling rate. A female French native speaker who was instructed to “speak
23
24 clearly” produced sequences of /aba/ and /apa/. Sixteen tokens were selected from a large
25
26 sample (around 150) in each phonetic category to be comparable in duration, intensity and
27
28 pitch (in order to reduce variability across categories). Mean duration was 651.5 ms for /aba/
29
30 (range: 592-692 ms; $SD = 36$ ms), and 630 ms for /apa/ (range: 570-702 ms; $SD = 47$ ms). All
31
32 stimuli were equated in global root-mean-square (RMS) level. The F_0 was estimated at 242
33
34 Hz using the YIN algorithm (de Cheveigné & Kawahara, 2002).
35
36
37

38 Four speech-processing conditions were used (the spectrograms of processed stimuli are
39
40 shown in figure 1)³. In the first condition (called “32-band AM+FM speech”), the original
41
42 speech signal was decomposed into 32, 1- ERB_N -wide frequency bands using zero-phase, 6th-
43
44 order Butterworth bandpass filters (36 dB/octave rolloff) with central frequency (CF) ranging
45
46
47

48 ³ A pilot experiment using an ABX, forced-choice discrimination task was initially conducted on 40 normal-
49
50 hearing adults using the intact and vocoded stimuli, in order to verify that the processed sounds can be
51
52 discriminated. This pilot experiment showed that, after a very short practice, normal-hearing adults discriminated
53
54 almost perfectly voiced and unvoiced consonants in each speech-processing condition (>90% correct
55
56 discrimination). This finding is consistent with the outcome of the original speech-identification experiment
57
58 conducted by Shannon *et al.* (1995), showing that voicing perception is nearly perfect in adults as long as AM
59
60 cues below 16 Hz are presented in at least four broad spectral regions. Interestingly, all listeners reported having
recognized the disyllables /aba/ and /apa/ except those presented with “4-band AM speech”. In this speech-
processing condition, stimuli were not recognized as speech sounds.

1
2
3
4
5 from 80 to 8,020 Hz. The Hilbert transform was then applied to each bandpass filtered speech
6 signal to extract the AM component and FM carrier. The AM component was low-pass
7 filtered using a zero-phase Butterworth filter (36 dB/octave rolloff) with a cutoff frequency
8 set to $ERB_N/2$. The final narrow-band speech signal was obtained by multiplying each sample
9 of the FM carrier by the filtered AM function. The narrow-band speech signals were finally
10 added up and the level of the wideband speech signal was adjusted to have the same RMS
11 value as the input signal. Thus, the vocoded speech signals retained the original AM and FM
12 speech cues within each of the 32 analysis frequency bands.
13
14
15
16
17
18
19
20
21
22

23 In the second condition (called “32-band AM speech”), the same signal processing
24 scheme was used as in the “32-band AM+FM speech” condition, except that the FM carrier
25 was replaced by a band of pink noise in each analysis frequency band. Thus, the resulting
26 vocoded speech signal retained AM speech cues within 32 bands, but discarded the original
27 (within-channel) FM speech cues.
28
29
30
31
32
33

34 In the third condition (called “32-band AM<16Hz speech”), the same signal
35 processing scheme was used as in the “32-band AM speech” condition, except that the AM
36 component was low-pass filtered with a cutoff frequency of 16 Hz for each of the 32 bands in
37 order to remove fast, F0-related AM cues. Thus, the resulting vocoded speech signal retained
38 mainly the slowest (<16 Hz) AM speech cues within 32 bands, and discarded the original FM
39 speech cues.
40
41
42
43
44
45
46

47 In the last condition (called “4-band AM speech”), the same signal processing scheme
48 was used as in the “32-band AM speech” condition, except that AM cues were extracted from
49 only 4, broad ($8-ERB_N$ wide) frequency bands. Thus, the original FM speech cues were
50 discarded, and AM cues were distorted substantially compared to the original AM speech
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 cues. This vocoder also reproduces the sound processing typically achieved by current
6 cochlear implant' sound processors (*cf.* Shannon *et al.*, 1995).
7
8

9
10 /Insert Figure 1 about here/
11

12 In each condition, 4 different sequences were created. Each sequence was composed of
13 4 tokens of the same phonetic category, repeated 4 times in a different random order. Thus, all
14 sequences have the same number of stimuli. Two sequences were used for the familiarization
15 phase. Two sequences were used for the test phase. The tokens used in the test phase for each
16 phonetic category were different from the ones used in the familiarization phase. The inter-
17 stimulus interval was varied randomly between 390 and 600 ms all along the 16-item
18 sequences. This random variation was introduced to prevent infants from using small
19 variations in duration between items within and between categories. Finally, all the sound
20 sequences were equated in duration (18 s).
21
22
23
24
25
26
27
28
29
30

31 Procedure

32 The experiment was conducted inside a sound-attenuated room where three lamps were
33 fixed: a green one on the center wall, and a red one on each side wall. Below the green lamp
34 was a hole for the lens of a video-camera. Out of the infant view, two loudspeakers were
35 placed behind the red lamps and delivered the speech stimuli at 70 dB SPL (RMS) at the level
36 of the infants' head. The infant was seated on the caregiver's lap in the center of booth. The
37 caregiver was instructed not to speak or interfere in any way with the infant's behavior and
38 wore ear plugs plus headphones delivering masking music.
39
40
41
42
43
44
45
46
47
48

49 The entire experimental session was controlled by a computer outside the booth, and a
50 TV screen was connected to the camera. The experimenter sat outside the booth, looked at the
51 video of the infant on the TV screen to monitor infant's looking behavior. The experimenter
52 used a response box composed of three buttons corresponding to the three lights inside the
53
54
55
56
57
58
59
60

1
2
3
4
5 booth. The response box was connected to the computer controlling the experiments. The
6
7 experimenter pressed the buttons of the response box according to the direction of the infant's
8
9 head (center, right or left). During the entire session, the experimenter wore headphones
10
11 delivering masking music and was unaware of the nature of the displayed sequences (familiar
12
13 or novel) during the test.

14
15
16 During the familiarization and the test phase, each experimental trial started by the
17
18 blinking of the green center lamp. When the infant oriented to the green lamp, the
19
20 experimenter pressed the "center" button, switching the green lamp off, and simultaneously
21
22 one of the red lamps on. When the infant turned his/her head towards the red blinking lamp,
23
24 the experimenter pressed the "right" or "left" button, triggering the presentation of the
25
26 auditory stimulus from the corresponding loudspeaker. The trial was stopped when the infant
27
28 turned his/her head away for more than 2 s, corresponding to the button released by the
29
30 experimenter, or when the end of the stimulus sequence was reached. Information about the
31
32 duration of the head-turn was automatically stored on the computer.

33
34
35 During the familiarization phase, infants heard two sequences of the same phonetic
36
37 category (half of the infants heard /aba/ and the other half /apa/). The familiarization trials
38
39 continued (as described above) until infants listened to each sequence for a cumulative
40
41 duration of 30 s. This was achieved to complete 60 s of familiarization time. The number of
42
43 trials necessary to reach this criterion differed across participants; those who reached the
44
45 criterion with a number of trials significantly larger than the group mean were not included in
46
47 the analyses.

48
49
50 In the test phase, two different sequences of the familiarized item, and two sequences of
51
52 the novel item were displayed twice in random order and counterbalanced between the left
53
54 and right side of presentation. Thus, the test phase was composed of 8 trials for each subject.
55
56
57
58
59
60

1
2
3
4
5 The order of test sequences was counterbalanced between subjects (half of the subjects
6 received a familiar stimulus sequence as first test trial while, the other half received a novel
7 sequence as first test trial).
8
9

10 11 **Results**

12
13
14 Mean looking times were calculated for each participant across the 8 test trials (4
15 novel and 4 familiar sequences).
16
17

18 An analysis of variance (ANOVA) for repeated measures was run with 4 Conditions as
19 the between-subject factor and 2 Types of sequences (Familiar *versus* Novel) as the within-
20 subject factor. This analysis did not indicate any main effect of Condition ($F(3,84) = 0.78$; $p =$
21 $.51$) or Type of sequence ($F(1,84) = 0.26$; $p = .61$), but a significant interaction between
22 Condition and Type of Sequence ($F(3,84) = 3.06$; $p = .033$). This interaction is mainly due to
23 the fact that in the “32-band AM+FM” condition, infants listened longer to the novel
24 sequences than to the familiar sequences (9.3 s; $SD = 2.7$ s *versus* 8.2 s; $SD = 3.2$ s,
25 respectively) compared to the “4-band AM” condition in which infants listened less to the
26 novel sequences than to the familiar (7.1 s; $SD = 2.6$ s *versus* 8.1 s; $SD = 2.1$ s, respectively).
27 In the other two conditions (“32-band AM” and “32-band AM<16 Hz”), the mean looking
28 times were similar for the novel and the familiar sequences (see figure 2).
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 Planned comparisons confirmed that the significant interaction mentioned above is
44 mainly due to the interaction between the “32-band AM+FM” and “4-band AM” conditions
45 and Type of sequences ($F(1,84) = 8.59$; $p = .004$). Additionally, paired *t-tests* were carried out
46 for each condition separately. A significant preference for the novel sequences was observed
47 in the “32-band AM+FM” condition ($t(21) = 2.33$, $p = .03$), and a significant preference for
48 the familiar sequences was observed in the “4-band AM” condition ($t(21) = 2.43$, $p = .02$). No
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 significant difference was observed in the other two “32-band AM” and “32-band AM<16
6
7 Hz” conditions ($p > .05$ in both conditions).
8

9
10 /Insert Figure 2 about here/
11

12 Discussion

13
14 As expected, infants exhibited significantly longer looking times for the novel stimuli in
15
16 the “32-band AM+FM” speech condition, indicating that they discriminated /aba/ from /apa/
17
18 when provided with intact AM and FM speech cues. This novelty preference is a classical
19
20 demonstration of infants’ discrimination capacities. However, the familiarity preference can
21
22 also be taken as an indication of discrimination (*e.g.*, Hunter & Ames, 1988). In behavioral
23
24 procedures including a familiarization phase, familiarization is supposed to bias stimulus
25
26 preference in such a way that all significant differences (whatever the direction, since
27
28 familiarized stimuli were counterbalanced across subjects) would indicate that infants have
29
30 processed novel and familiar stimuli differently. Here, an unexpected preference for familiar
31
32 stimuli was observed for 6-month-old infants in the “4-band AM” speech condition, where the
33
34 frequency resolution of the vocoder was reduced to 4 broad (*i.e.*, 8-ERB_N-wide) frequency
35
36 bands. Although there are a number of hypotheses as to which factors affect the direction of
37
38 preferences, no consensus has emerged (*e.g.*, Hunter & Ames, 1988; Rose, Gottfried, Melloy-
39
40 Carminar, & Bridger, 1982; Thiessen & Saffran, 2003; Wagner & Sakovits, 1986). Hunter
41
42 and Ames (1988) proposed a model of infant preferences for novel and familiar stimuli based
43
44 on the interaction among three factors, namely age, familiarization time and task difficulty.
45
46 According to this model, the present results in the “4-band AM” speech condition may be
47
48 explained by the difficulty to process this relatively impoverished signal. In this speech
49
50 condition, the original FM speech cues were discarded, and AM cues were distorted
51
52 substantially compared to the “32-band AM+FM” condition and to the AM speech cues that
53
54
55
56
57
58
59
60

1
2
3
4
5 listeners typically extract from the relatively narrow (1-ERB_N-wide) frequency bands
6
7 corresponding to the outputs of their cochlear filters (Kates, 2011).
8

9
10 The results obtained in the “32-band AM” condition were even more unexpected given
11 that the stimuli used here were supposed to convey more accurate modulation information
12 than the “16-band AM” speech stimuli previously used by Bertoncini *et al.* (2011). Indeed, in
13 the present study, a higher frequency resolution was used and higher AM rates (*e.g.*, periodic
14 AM fluctuations at F0) were transmitted in high-CF analysis bands. Here, the AM component
15 was lowpass filtered at ERB_N/2 within each band, *versus* 64 Hz in Bertoncini *et al.* (2011). It
16 is thus possible that the inherent random amplitude fluctuations of the noise carriers masked
17 to some extent the original AM speech cues (*e.g.*, Dau, Kollmeier, & Kohlrausch, 1997a,b;
18 Dau, Verhey, & Kohlrausch, 1999; Lorenzi *et al.*, 2001; see also Kates, 2011). Such a
19 “modulation masking” effect did not occur in Bertoncini *et al.* (2011)’ study because pure-
20 tone carriers were used instead of band of noise. Alternatively, another factor could have
21 accounted for this difference. As mentioned above, the preset duration of the familiarization
22 period could have been inappropriate (*i.e.*, too short) when artificially degraded stimuli such
23 as noise-vocoded speech sounds were used. In Bertoncini *et al.*’s (2011), 6-month-old infants
24 received no familiarization, and were found to prefer spontaneously alternating series of tone-
25 vocoded speech sounds to repeating ones. In the present version of the procedure,
26 discrimination could require further perceptual processing than in the previous study, where
27 the repetitive juxtaposition of /aba/ and /apa/ could have prompted immediate comparison. It
28 may be the case that the information needed to represent a given stimulus category has to be
29 gathered and correctly formatted before being compared to the (correctly formatted) other
30 category. This is consistent with the studies conducted by Holt and Carney (2005, 2007) on
31 children and adults. These authors proposed that “robustness” of the internal representation of
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 speech stimuli and correct discrimination of a given speech contrast increase with increasing
6 repetition/presentation of the stimuli. Again, the duration of the familiarization phase used in
7 Experiment 1 might have been too short for infants to complete such a perceptual comparison,
8 and to retrieve the original AM speech cues from the statistical amplitude fluctuations
9 produced by noise carriers. Hunter and Ames (1988) indicated that the absence of any
10 preference following a familiarization phase can be explained by a too short familiarization
11 time. The duration necessary to be properly familiarized with a given stimulus may depend on
12 the vocoded condition. The high variability in infants' responses observed in the "32-band
13 AM" and "32-band AM<16Hz" speech conditions may also question the effectiveness of the
14 familiarization period, suggesting a transitional phase in preference in these 2 conditions (see
15 also Hunter & Ames, 1988; Rose *et al.*, 1982).
16
17
18
19
20
21
22
23
24
25
26
27
28

29 The absence of a significant difference between the "32-band AM+FM", "32-band
30 AM" and "32-band AM<16Hz" speech conditions precluded any conclusion about the effects
31 of degrading AM and/or FM information. Thus, when the AM and FM speech cues are
32 severely degraded, infants may need an extended exposure with the contrasted stimuli. In our
33 view, this could also result in a more consistent effect of preference for one type of sequence.
34 A second experiment in which the familiarization time was extended to a *minimum* of 2 min
35 was conducted to test this hypothesis.
36
37
38
39
40
41
42
43
44
45
46

47 **Experiment 2:**

48
49 Four groups of twenty 6-month-old French-learning infants were tested with a
50 minimum familiarization duration extended to 2 min.
51
52

53 **Participants**

54
55
56
57
58
59
60

1
2
3
4
5 Data from 80 infants were analyzed (45 boys and 35 girls range: 5 months 23 days; 7
6 months 3 days; mean = 6 months and 13 days; $SD = 8$ days). The data from 18 additional
7 infants were not included for the following reasons: fussing and crying ($n = 14$), looking time
8 lower than 1500 ms for one trial ($n = 1$), outlier number of familiarization trials ($n = 2$) and
9 outlier looking time differences between novel and familiar conditions ($n = 1$).
10
11
12
13
14
15

16 **Stimuli**

17
18 Infants were presented with the same stimuli and sequences of stimuli as those used in
19 Experiment 1.
20
21

22 **Procedure**

23
24 The same procedure (HPP) was used, except that the familiarization time criterion was
25 changed. In the familiarization phase, infants heard two sequences of the same phonetic
26 category, and they had to listen to each sequence for at least 60 s; therefore, the
27 familiarization phase had a minimum cumulative duration of 120 s. The test phase was left
28 unchanged: each subject received 8 trials containing two different sequences of new tokens of
29 the familiarized item, and two sequences of the novel item, that were played twice in a
30 random order.
31
32
33
34
35
36
37
38
39
40

41 **Results**

42
43 As for Experiment 1, the mean looking times for novel and familiar stimuli were
44 calculated for each participant on the 8 trials of the test phase. An ANOVA for repeated
45 measures was run with 4 Conditions as the between-subject factor and 2 Types of sequences
46 (Familiar *versus* Novel) as the within-subject factor. The main effect of Condition was not
47 significant ($F(3,76) = 1.29$; $p = .28$). However, the analysis showed a significant effect of
48 Type of sequence ($F(1,76) = 8.88$; $p = .004$) and a significant interaction between Condition
49
50
51
52
53
54
55
56
57
58
59
60

and Type of sequence ($F(3,76) = 5.67; p = .0015$). Planned comparisons indicated that only the condition “32-band AM< 16Hz” differed significantly from the three others ($\alpha = .05$).

Once again, separate analyses revealed some differences between conditions. A significant preference for the novel sequences compared to the familiar ones was observed in three conditions: in the “32-band AM+FM” condition (7.29 s; $SD = 1.7$ s versus 6.08 s; $SD = 2.07$ s, respectively; $t(19) = 2.22, p = .039$), in the “32-band AM” condition (6.2 s, $SD = 2.6$ s versus 4.8 s, $SD = 2.1$ s, respectively, ($t(19) = 2.99, p = .014$), and in the “4-band AM” condition (7.34 s; $SD = 3.2$ s versus 5.9 s; $SD = 2.5$ s, respectively; $t(19) = 2.50, p = .022$). In the “32-band AM<16 Hz” condition, the mean looking time for novel sequences was shorter than for familiar sequences (6.3 s, $SD = 3.1$ s versus 7.3 s, $SD = 3.1$ s). This difference was significant ($t(19) = -2.55; p = .028$), indicating that infants listened to familiar stimuli for a longer time in this condition (see figure 3).

/Insert Figure 3 about here/

Discussion

In this second experiment, each condition was re-tested with a preset familiarization time of 2 min instead of 1 min (as used in Experiment 1). With a longer familiarization time, the 6-month-old infants showed the classical pattern of preference for novelty in three conditions out of four. However, infants showed a preference for familiar stimuli in the “32-band AM<16 Hz” condition, that is when the fastest (> 16 Hz) AM fluctuations related to F0 variations were reduced.

The results of this second experiment revealed that discrimination of /aba/ and /apa/ was possible, and indicated by a classical novelty preference in both “32-band AM” and “4-band AM” conditions. The preference pattern in the “32-band AM<16 Hz” condition turns out

1
2
3
4
5 to be significant, and confirm the hypothesis that extended familiarization is necessary to
6
7 evidence discrimination responses in some vocoder conditions.
8
9

10 11 12 **General discussion**

13
14 The present study was designed to explore how 6-month-old infants process a French
15 voicing contrast on the basis of AM and FM cues, at an age when their perceptual
16 mechanisms are not completely tuned to their native language. The disyllables (/aba/ versus
17 /apa/) were used and several speech-processing conditions were designed to investigate
18 whether infants discriminate this contrast when the modulation properties of speech sounds
19 are severely degraded.
20
21
22
23
24
25
26

27
28 *Discrimination of speech modulation cues.* Six-month-old infants were found to
29 discriminate the voicing contrast with the vocoded speech stimuli.
30
31

32 In the second experiment, results demonstrated that within-channel FM cues are not
33 necessary to discriminate voicing at 6 months of age (*cf.* “32-band AM” condition). They also
34 demonstrated that infants could discriminate voicing in the absence of periodic AM
35 fluctuations related to F0 (*cf.* “32-band AM<16 Hz” condition). However, even with an
36 extended familiarization time, infants showed an opposite pattern of discrimination compared
37 to the “32-band AM” condition where fast AM variations related to F0 were preserved. This
38 suggests that infants are sensitive to a degradation of fast F0 cues in the AM domain, an
39 indication that, as for adults (*e.g.*, Rosen, 1992), the presence of fast, F0-related periodic
40 information constitutes reliable cues to voicing cues for infants. This is also consistent with
41 the outcome of previous psychophysical studies showing that the ability to detect changes in
42 AM cues as a function of time – *i.e.*, auditory temporal resolution – is efficient by 6 months of
43 age (Levi & Werner, 1996; for a review, see Saffran *et al.*, 2006 and Werner & Gray, 1998).
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 In Experiments 1 and 2, the results showed that infants are able to discriminate speech
6 signals containing only AM cues extracted within a small number of frequency bands (“4-
7 band AM” condition). Moreover, in Experiment 1, the preference pattern for familiar stimuli
8 observed in this condition compared to the “32-band AM+FM” and “32-band AM” conditions
9 suggests that infants are sensitive to a reduction in frequency resolution. This is consistent
10 with the demonstration of adult-like frequency selectivity in infants by the age of 6 months
11 (*e.g.*, Abdala & Folsom, 1995; Spetner & Olsho, 1990).
12
13
14
15
16
17
18
19

20 *The importance of auditory exposure to the stimuli.* These findings suggest that at this
21 early age, auditory processes are capable of making subtle speech distinctions despite severe
22 distortions of speech modulation cues. Nevertheless, it is important to keep in mind that the
23 demonstration of discrimination in infants was not straightforward in all conditions of
24 degradation. In the “32-band AM+FM” condition, increasing the familiarization time did not
25 affect the infants’ pattern of responses. However, in the “4-band AM” condition, the increase
26 in familiarization time by 1 min affected the infants’ response pattern, which changed from a
27 preference to familiar stimuli to a preference to novel stimuli. In the “32-band AM” condition,
28 the extended auditory exposure led to a preference for the novel stimuli. In the “32-band AM<
29 16Hz” condition, infants showed a significant preference for familiar sequences. Thus, a pre-
30 set and relatively short time of exposure may not have allowed the 6-month-old infants to
31 fully process speech cues in each vocoder condition, and build up the detailed representations
32 of the vocoded signals required for robust discrimination (*cf.* Holt, 2011).
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 This suggests that the present version of HPP based on a fixed familiarization time
50 period might not have optimally assessed discrimination capacities of degraded stimuli (and
51 quite unfamiliar speech sounds) in infants. The familiarization phase was aimed to provide an
52 equivalent amount of auditory experience with a particular sound to a group of infants.
53
54
55
56
57
58
59
60

1
2
3
4
5 However, it is important to keep in mind that vocoded stimuli strongly differed in terms of
6 amount of spectro-temporal degradation. A procedure using an infant controlled habituation
7 time would be more efficient to assess the discrimination capacity with such degraded speech
8 stimuli. Future studies are warranted to assess whether or not familiarization time should be
9 adapted to each infant in order to reduce inter-individual variability across the different
10 conditions of degradation.
11
12
13
14
15
16
17

18 *Different response patterns of infants.* A non-classical preference pattern was observed
19 for infants when AM and FM cues were simultaneously degraded as in the “32-band<16 Hz”
20 and “4-band AM” conditions (Experiment 2 and Experiment 1, respectively). In these two
21 conditions, infants showed a preference for the familiar stimuli instead of the more typical
22 preference for novel stimuli observed in the “32-band AM+FM” and “32-band AM”
23 conditions. Altogether, these results are congruent with the assumption that a familiarity
24 preference could be related to some difficulties in processing stimuli (Hunter & Ames, 1988).
25 This difference in response patterns suggests that the mechanisms involved in voicing
26 perception may differ according to the severity of modulation distortions, and the subsequent
27 “difficulty” in processing these highly distorted – and thus, unfamiliar - signals (Hunter &
28 Ames, 1988; Rose *et al.*, 1982).
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 This suggestion is consistent with the fact that adult participants tested in a pilot study
44 (see footnote 2) could not readily recognize the linguistic nature of the vocoded stimuli in the
45 “4-band AM” condition, and required extensive training to identify consonants, vowels or
46 sentences in the same speech-processing condition as previously shown by Shannon *et al.*
47 (1995). This suggestion is also consistent with the notion that part of the processing
48 mechanisms recruited for a given task may change whether adults are informed (or not) about
49 the nature of what they are supposed to listen to (see Liebenthal, Binder, Piorkowski, &
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 Remez, 2003). When the nature of the stimuli is unknown and unfamiliar, perceptual
6 mechanisms may differ from those activated by an intact speech signal.
7
8

9
10 In Experiment 2, the particular pattern of results observed in the “32-band AM<
11 16Hz” condition may reveal the importance of the fast AM cues related to F0 fluctuations for
12 voicing discrimination at 6 months.
13
14

15
16 *Clinical implications.* In the “4-band AM” condition, six-month-old infants (with
17 normal hearing) were tested with speech stimuli processed by a vocoder simulating the sound
18 processing typically achieved by current cochlear implant sound processors (e.g., Friesen,
19 Shannon, Baskent, & Wang, 2001; Shannon *et al.*, 1995). The results showed that infants
20 could still discriminate voicing despite important degradations in both AM and FM cues.
21 However, a classical “novelty” discrimination response seems to require an extended
22 exposure with the contrasted stimuli when the speech modulation cues were severely
23 degraded by filtering or spectral reduction.
24
25
26
27
28
29
30
31
32
33

34 The implications of the present results for the understanding of speech perception
35 capacities for infants wearing a cochlear implant should however be treated with caution.
36 Indeed, the present results were obtained using degraded speech with normal-hearing infants,
37 as well as a single speech contrast in a single context. Still, these results are very encouraging
38 for understanding speech development in deaf infants wearing cochlear implants, and point to
39 the necessity to investigate the basic auditory abilities needed to benefit from very early
40 cochlear implantation (e.g., Holt & Svirsky, 2008; Miyamoto, Houston, Kirk, Perdew, &
41 Svirsky, 2003; Nikolopoulos, Archbold, & O’Donoghue, 1999; Svirsky, Robbins, Kirk,
42 Pisoni, & Myamoto, 2000). Further data will be needed to understand how the degraded
43 signals delivered by cochlear implant processors are processed throughout the development of
44 speech perception. In addition to their clinical relevance, these results might draw attention to
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 the role of basic auditory processes upon early speech perception before or during the
6 development of language-specific tuning (Kuhl *et al.*, 2008).
7
8
9

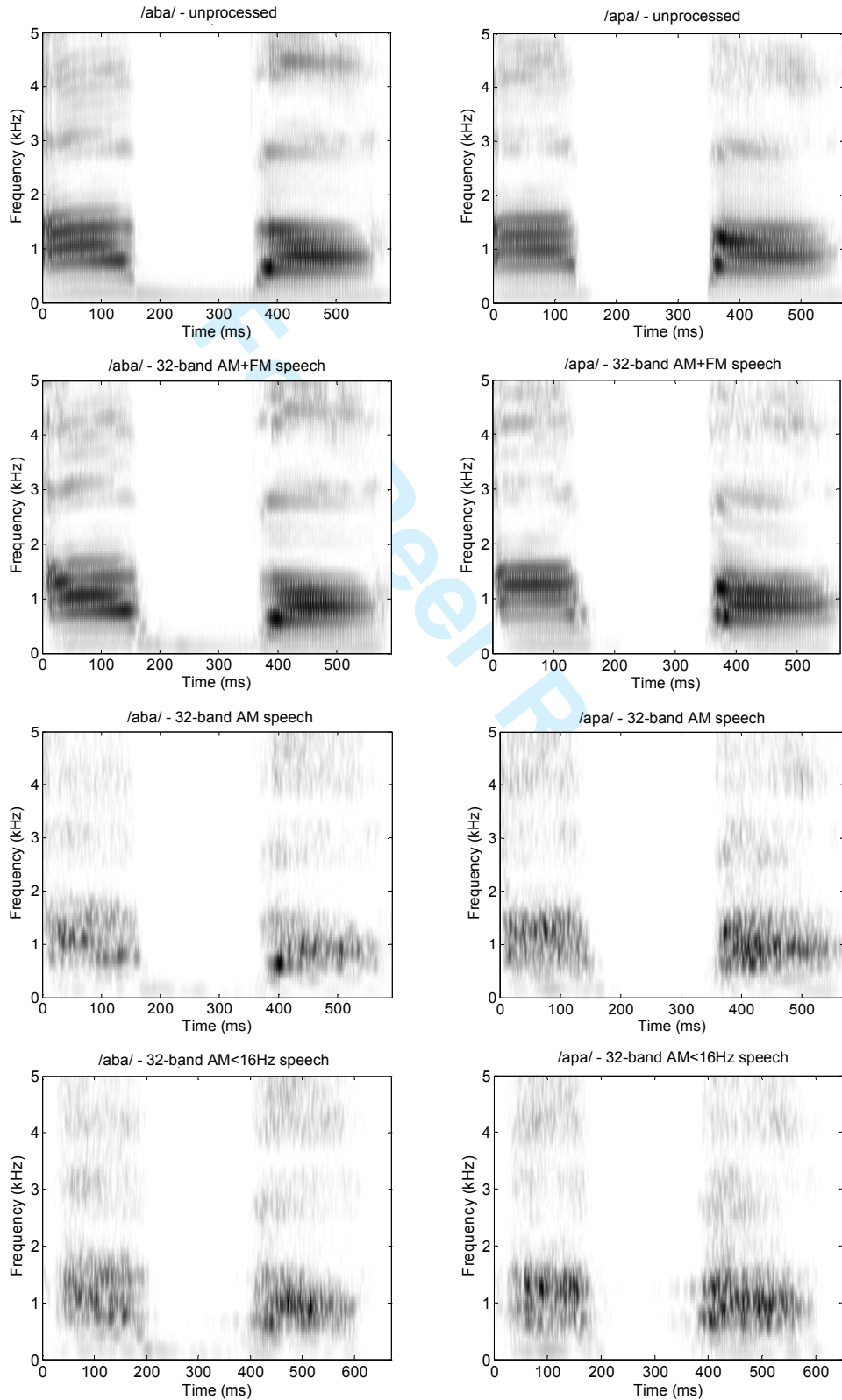
10 11 12 **Conclusion**

13
14 The present study explored how infants process the acoustic information related to
15 phonetic differences using a novel approach stressing the importance of modulation speech
16 cues. The results indicate that the relatively fast AM and FM cues in speech are not necessary
17 for French-learning 6-month-old infants to discriminate a voicing contrast in silence. These
18 results demonstrate that the perception of voicing is robust as early as 6 months of age.
19 However, the duration of initial auditory exposure (*i.e.*, familiarization time) to the vocoded
20 stimuli had to be increased to evidence discrimination abilities in all speech-processing
21 conditions. Finally, the results show that normal-hearing infants can use the impoverished
22 speech modulation cues delivered by prosthetic devices such as cochlear implants to
23 discriminate voicing, provided they are sufficiently exposed and familiarized to these
24 degraded speech stimuli.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 **Acknowledgments**

44 This work was supported by ANR and program “Investissement d’avenir”(ANR-11-IDEX-
45 0001-02-PSL; Labex “Institut d’Etude de la Cognition”). C. Lorenzi was supported by a grant
46 (HEARFIN Project) from ANR. The authors would like to thank warmly the parents for their
47 participation in this study.
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

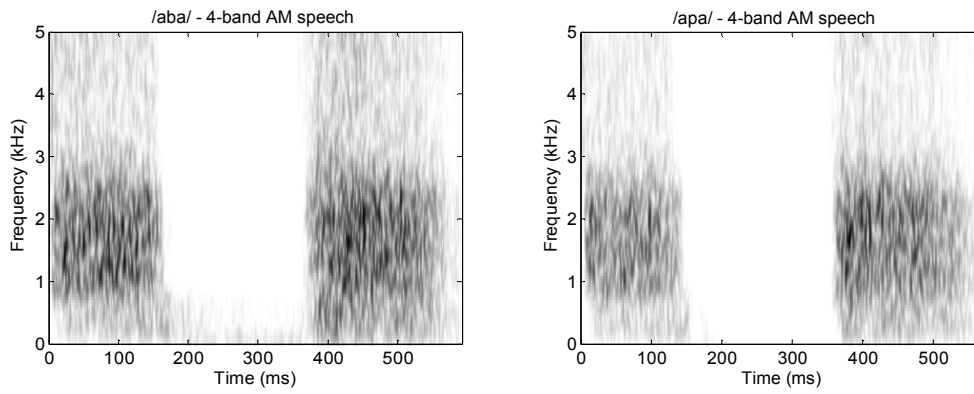


Figure 1 Spectrograms of /aba/ (left panels) and /apa/ (right panels). From top to bottom panels: unprocessed, “32-band AM+FM”, “32-band AM”, “32-band AM<16Hz”, and “4-band AM” speech conditions.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

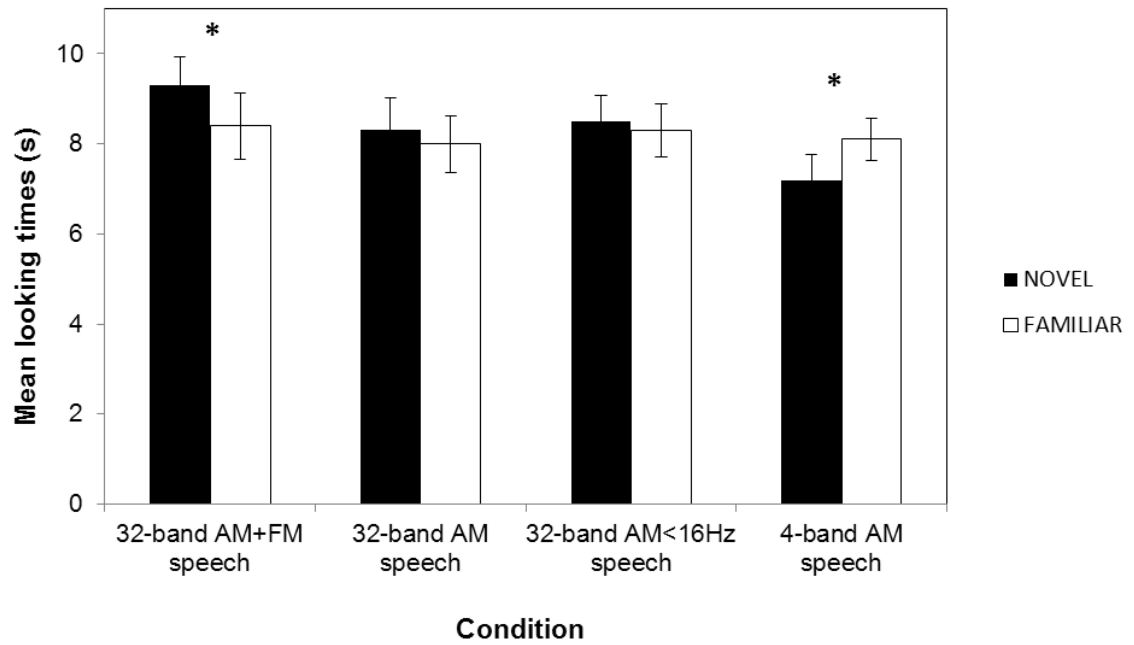


Figure 2 Mean looking times in Experiment 1 for familiar and novel stimuli during the test phase for each speech-processing condition (errors bars represent the standard errors).

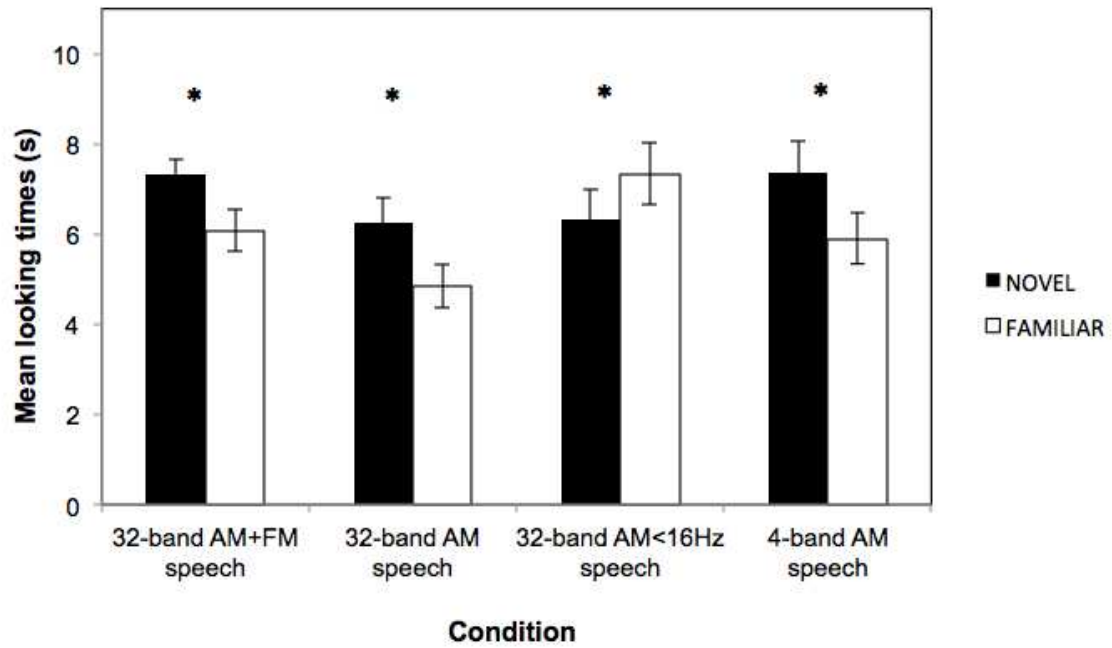


Figure 3 Mean looking times in Experiment 2 for familiar and novel stimuli during the test phase for each speech-processing condition (errors bars represent the standard errors).

References

- Abdala, C., & Folsom, R.C. (1995). The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking. *Journal of the Acoustical Society of America*, *98*, 921-930.
- Ardoint, M., & Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing research*, *260*, 89-95.
- Aslin, R.N. (1989). Discrimination of frequency transitions by human infants. *Journal of the Acoustical Society of America*, *86*, 582-590.
- Bertoncini, J., Nazzi, T., Cabrera, L., & Lorenzi, C. (2011). Six-month-old infants discriminate voicing on the basis of temporal envelope cues. *The Journal of the Acoustical Society of America*, *129*, 2761-2764.
- Bertoncini, J., Serniclaes, W., & Lorenzi, C. (2009). Discrimination of speech sounds based upon temporal envelope versus fine structure cues in 5-to-7 year-old children. *Journal of Speech Language and Hearing Research*, *52*, 682-695.
- Bijeljac-Babic, R., Serres, J., Hohle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS ONE*, *7*(2), e30843.
- Bosch, L., & Sebastian-Galles, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29-49.
- Brugge, J.F., Javel, E., & Kitzes, L.M. (1978). Signs of functional maturation of peripheral auditory system in discharge patterns of neurons in anteroventral cochlear nucleus of kitten. *Journal of Neurophysiology*, *41*, 1557-1559.
- Buss, E., Hall, J.W.III, Grose, J.H., & Dev, M.B. (1999). Development of adult-like performance in backward, simultaneous, and forward masking. *Journal of Speech Language and Hearing Research*, *42*, 844-849.
- Colombo, J., & Horowitz, F.D. (1986). Infants attentional responses to frequency modulated sweeps. *Child Development*, *57*, 287-291.
- Dau, T., Kollmeier, B., & Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation : I. Modulation detection and masking with narrow-band carriers. *Journal of the Acoustical Society of America*, *102*, 2892-2905.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Dau, T., Kollmeier, B., & Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation : II. Spectral and temporal integration in modulation detection. *Journal of the Acoustical Society of America*, *102*, 2906-2919.
- Dau, T., Verhey, J., & Kohlrausch, A. (1999). Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band carriers. *Journal of the Acoustical Society of America*, *106*, 2752-2760.
- de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, *111*, 1917-1930.
- Deng L., & Geisler, C.D. (1987). Responses of auditory-nerve fibers to nasal consonant-vowel syllables. *Journal of the Acoustical Society of America*, *82*, 1977-1988.
- Diedler, J., Pietz, J., Bast, T., & Rupp (2007). Auditory temporal resolution in children assessed by magnetoencephalography. *Neuroreport*, *18*, 1691-1695.
- Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *Journal of the Acoustical Society of America*, *97*, 585-592.
- Dudley, H. (1939). Remaking speech. *The Journal of the Acoustical Society of America*, *11*, 165-165.
- Eaves, J.M., Summerfield, A.Q., & Kitterick, P.T. (2011). Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes. *Journal of the Acoustical Society of America*, *130*, 501-507.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infant. *Science*, *171*, 303-306.
- Eisenberg, L.S., Shannon, R.V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*, *107*, 2704-2710.
- Friesen, L.M., Shannon, R.V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America*, *110*, 1150-1163.
- Gilbert, G., Bergeras, I., Voillery, D., & Lorenzi, C. (2007). Effects of periodic interruption on the intelligibility of speech based on temporal fine-structure or envelope cues. *Journal of the Acoustical Society of America*, *122*, 1336-1339.
- Gilbert, G., & Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *Journal of the Acoustical Society of America*, *119*, 2438-2444.

- 1
2
3
4
5 Glasberg, B.R., & Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-
6 noise data. *Hearing Research*, 47, 103-138.
- 7
8 Gnansia, D., Pean, V., Meyer, B., & Lorenzi, C. (2009). Effects of spectral smearing and
9 temporal fine structure degradation on speech masking release. *Journal of the*
10 *Acoustical Society of America*, 125, 4023-4033.
- 11
12 Hall, J.W.III, & Grose, J.H. (1994). Development of temporal resolution in children as
13 measured by the temporal modulation transfer function. *Journal of the Acoustical*
14 *Society of America*, 96, 150-154.
- 15
16 Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Wright, K., Druss, B., & Kennedy, L.
17 (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269-286.
- 18
19 Hhle, B., Bijeljic-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). The
20 development of language specific prosodic preferences during the first half year of
21 life: evidence from German and French. *Infant Behavior and Development*, 32, 262-
22 274.
- 23
24 Holt, R.F. (2011). Enhancing speech discrimination through stimulus repetition. *Journal of*
25 *Speech, Language, and Hearing Research*, 54, 1431-1447.
- 26
27 Holt, R.F., & Carney, A.E. (2005). Multiple looks in speech sound discrimination in adults.
28 *Journal of Speech, Language, and Hearing Research*, 48, 922-943.
- 29
30 Holt, R.F., & Carney, A.E. (2007). Developmental effects of multiple looks in speech sound
31 discrimination. *Journal of Speech, Language, and Hearing Research*, 50, 1404-1424.
- 32
33 Holt, R.F., & Svirsky, M.A. (2008). An exploratory look at paediatric cochlear implantation:
34 Is earliest always best? *Ear and Hearing*, 29, 492-511.
- 35
36 Hopkins, K., Moore, B.C.J., & Stone, M.A. (2008). Effects of moderate cochlear hearing loss
37 on the ability to benefit from temporal fine structure information in speech. *Journal of*
38 *the Acoustical Society of America*, 123, 1140-1153.
- 39
40 Hopkins, K., Moore, B.C.J., & Stone, M.A. (2010). The effects of the addition of low-level,
41 low-noise noise on the intelligibility of sentences processed to remove temporal
42 envelope information. *Journal of the Acoustical Society of America*, 128, 2150-2161.
- 43
44 Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2009).
45 French native speakers in the making: from language-general to language-specific
46 voicing boundaries. *Journal of Experimental Child Psychology*, 104, 353-366.
- 47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5 Hunter, M.A., & Ames, E.W. (1988). A multifactor model of infant preferences for novel and
6 familiar stimuli. *Advances in infancy research*, 5, 69-95.
7
8 Johnson, D.H. (1980). The relationship between spike rate and synchrony in responses of
9 auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America*, 68,
10 1115-1122.
11
12 Joris, P.X., & Yin, T.C. (1992). Responses to amplitude-modulated tones in the auditory
13 nerve of the cat. *Journal of the Acoustical Society of America*, 91, 215-232.
14
15 Joris, P.X., Schreiner, C.E., & Rees, A. (2004). Neural processing of amplitude-modulated
16 sounds. *Physiological Reviews*, 84, 541-577.
17
18 Kale, S., & Heinz, M.G. (2010). Envelope coding in auditory nerve fibers following noise-
19 induced hearing loss. *Journal of the Association for Research in Otolaryngology*, 11,
20 657-673.
21
22 Kates, J.M. (2011). Spectro-temporal envelope changes caused by temporal fine-
23 structure modification. *Journal of the Acoustical Society of America*, 129, 3981-3990.
24
25 Kiang, N.Y., Pfeiffer, R.R., Warr, W.B., & Backus, A.S. (1965). Stimulus Coding in the
26 Cochlear Nucleus. *Annals of Otolaryngology, Rhinology and Laryngology*, 74, 463-485.
27
28 Kettner, R.E., Feng, J.Z., & Brugge, J.F. (1985). Postnatal development of the phase-locked
29 response to low frequency tones of the auditory nerve fibers in the cat. *Journal of*
30 *Neuroscience*, 5, 275-283.
31
32 Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews.*
33 *Neuroscience*, 5, 831-43.
34
35 Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, S., Rivera-Gaxiola, M., & Nelson, T.
36 (2008). Phonetic learning as a pathway to language: new data and native language
37 magnet theory expanded (NLM-e). *Philosophical Transaction of the Royal Society of*
38 *London B Biological Science*, 363, 979-1000.
39
40 Kuhl P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants
41 show a facilitation effect for native language phonetic perception between 6 and 12
42 months. *Developmental Science*, 9, F13-F21.
43
44 Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic
45 experience alters phonetic perception in infant by 6 months of age. *Science*, 255, 606-
46 608.
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5 Levi, E.C., Folsom, R.C., & Dobie, R.A. (1995). Coherence analysis of envelope-following
6 responses (EFRs) and frequency-following responses (FFRs) in infants and adults.
7 *Hearing Research, 89*, 21-7.
8
9
10 Levi, E. C., & Werner, L. A. (1996). Amplitude modulation detection in infancy: Update on
11 3-month-olds. *Abstracts of the Association for Research in Otolaryngology, 19*, 142.
12
13 Liebenthal, E., Binder, J.R., Piorkowski, R.L., & Remez, R.E. (2003). Short-term
14 reorganization of auditory analysis induced by phonetic experience. *Journal of*
15 *cognitive neuroscience, 15*, 549-58.
16
17
18 Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B.C.J. (2006). Speech perception
19 problems of the hearing impaired reflect inability to use temporal fine structure.
20 *Proceedings of the National Academy of Sciences USA, 103*, 18866–18869.
21
22
23 Lorenzi, C., Simpson, M.I.G., Millman, R.E., Griffiths, T.D., Woods, W.P., Rees, A., &
24 Green, G.G.R. (2001). Second-order modulation detection thresholds for pure-tone
25 and narrow-band noise carriers. *Journal of the Acoustical Society of America, 110*,
26 2470-2478.
27
28
29
30 Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of
31 lexical tone perception in the first year of life. *Cognition, 106*, 1367-1381.
32
33
34 Miyamoto, R.T., Houston, D.M., Kirk, K.I., Perdew, A.E., & Svirsky, M.A. (2003) Language
35 Development in Deaf Infants Following Cochlear Implantation. *Acta Oto-*
36 *Laryngologica, 123*, 241-244.
37
38
39 Moore, B.C.J. (2003). *An introduction to the psychology of hearing, 5th Ed.* (Academic Press,
40 San Diego).
41
42 Moore, B.C.J., & Sek, A. (1996). Detection of frequency modulation at low modulation rates:
43 Evidence for a mechanism based on phase locking. *Journal of the Acoustical Society*
44 *of America, 100*, 2320-2331.
45
46
47 Moore, D.R., Cowan, J.A., Riley, A.; Edmondson-Jones, A.M., & Ferguson, M.A. (2011).
48 Development of Auditory Processing in 6- to 11-Yr-Old Children. *Ear & Hearing, 32*,
49 269-285.
50
51
52 Nazzi, T., Jusczyck, P.W., & Johnson, E.K. (2000). Language discrimination by English-
53 learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and*
54 *Language, 43(1)*, 1-19.
55
56
57
58
59
60

- 1
2
3
4
5 Nelson, P.B., Jin, S.H., Carney, A.E., & Nelson, D.A. (2003). Understanding speech in
6 modulated interference: cochlear implant users and normal-hearing listeners. *Journal*
7 *of the Acoustical Society of America*, *113*, 961-969.
8
9
10 Nikolopoulos, T.P., Archbold, S.M., & O'Donoghue, G.M. (1999). The development of
11 auditory perception in children following cochlear implantation. *International Journal*
12 *of Pediatric Otorhinolaryngology*, *49*, 189-191.
13
14
15 Olsho, L.W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior*
16 *Development*, *8*, 371-384.
17
18 Palmer, A.R., & Russel, I.J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and
19 its relation to the receptor potential of inner hair-cells. *Hearing research*, *24*, 1-15.
20
21 Plantinga, J., & Trainor, L.J. (2009). Melody recognition by two-month-old infants. *Journal*
22 *of the Acoustical Society of America*, *125*(2), EL58-EL62.
23
24
25 Qin, M.K., & Oxenham, A.J. (2003). Effects of simulated cochlear-implant processing on
26 speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*,
27 *114*, 446-54.
28
29
30 Rose, J.E., Brugge, J.F., Anderson, D.J., & Hind, J.E. (1967). Phase-locked response to low-
31 frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of*
32 *Neurophysiology*, *30*, 769-793.
33
34
35 Rose, S.A., Gottfried, A.W., Melloy-Carminar, P., & Bridger, W.H. (1982). Familiarity and
36 novelty preferences in infant recognition memory: Implications for information
37 processing. *Developmental Psychology*, *18*(5), 704-713.
38
39
40 Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects.
41 *Philosophical Transaction: Biological Sciences*, *336*, 367-373.
42
43
44 Saberi, K., & Hafter, E.R. (1995). A common neural code for frequency- and amplitude-
45 modulated sounds. *Nature*, *374*, 537-539.
46
47 Saffran, J., Werker, J., & Werner, L.A. (2006). The infant's auditory world: Hearing, speech
48 and the beginnings of language. In D. Kuhn & R.S. Siegler (ed.) *Handbook of Child*
49 *Psychology, Vol. 2, Cognition, Perception and Language* (6th edition) New York:
50 Wiley (pp. 58-108).
51
52
53 Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., & Zkelid, M. (1995). Speech
54 recognition with primarily temporal cues. *Science*, *270*, 303-304.
55
56
57
58
59
60

- 1
2
3
4
5 Sheft, S., Ardoit, M., & Lorenzi, C. (2008). Speech identification based on temporal fine
6 structure cues. *Journal of the Acoustical Society of America*, *124*, 562-575.
7
8 Sinex, D.G., & Geisler, C.D. (1983). Responses of auditory-nerve fibers to consonant-vowel
9 syllables. *Journal of the Acoustical Society of America*, *73*, 602-615.
10
11 Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., et al. (2009) Language specific
12 stress perception by nine-month-old French and Spanish infants. *Developmental*
13 *Science*, *12*, 914–919.
14
15 Smith, N.A., Trainor, L.J., & Shore, D.I. (2006). The development of temporal resolution:
16 between-channel gap detection in infant and adult. *Journal of Speech, Language, and*
17 *Hearing Research*, *49*, 1104-1113.
18
19 Smith, Z.M., Delgutte, B., & Oxenham, A.J. (2002). Chimaeric sounds reveal dichotomies in
20 auditory perception. *Nature*, *416*, 87-90.
21
22 Spetner, N.B., & Olsho, L.W. (1990). Auditory Frequency Resolution in Human Infancy.
23 *Child Development*, *61*, 632-652.
24
25 Svirsky, M.A., Robbins, A.M., Kirk, K.I., Pisoni, D.B., & Myamoto, R.T. (2000). Language
26 development in profoundly deaf children with cochlear implant. *Psychological*
27 *Science*, *11*, 153-158.
28
29 Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: use of stress and statistical cues to
30 word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*, 706–
31 716.
32
33 Trainor, J.L., Samuel, S.S., Desjardins, R.N., & Sonnadara, R.R. (2001). Measuring temporal
34 resolution in infants using mismatch negativity. *Neuroreport*, *12*, 2443-2448.
35
36 Trehub, S.E., Schneider, B.A., & Henderson, J.L. (1995). Gap detection in infants, children,
37 and adults. *Journal of Acoustical Society of America*, *98*, 2532-2541.
38
39 Wagner, S., & Sakovits, J. (1986). A process analyses of infant visual and cross-modal
40 recognition memory: Implications for an amodal code. In L. Lipsitt, & C. Rovee-
41 Collier, *Advances in infancy research*, *4*, 196–217.
42
43 Werker, J.F., & Tees, R.C. (1999). Influences on infant speech processing: toward a new
44 synthesis. *Annual Review of Psychology*, *50*, 509-35.
45
46 Werker, J.F., & Tees, R.C. (1983). Developmental changes across childhood in the perception
47 of non-native speech sounds. *Canadian Journal of Psychology*, *37*, 278-86.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5 Werner, L.A., Marean, G.C., Halpin, C.F., Spetner, N.B., & Gillenwater, J.M. (1992). Infant
6 auditory temporal acuity: gap detection. *Child Development, 63*, 260-272.
7
8 Werner, L.A., & Gray, L. (1998). Behavioral studies of hearing development. In E.W Rubel,
9 A.N.Popper, and R.R. Fay (Eds.). *Development of the auditory system*. Vol. 5,
10 Springer handbook of auditory research. New York: Springer-Verlag, 12-79.
11
12 Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D. (1989). Temporal resolution
13 in children. *Child Development, 60*, 611-624.
14
15 Zeng, F.G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.Y., & Chen, H. (2004). On
16 the dichotomy in auditory perception between temporal envelope and fine structure
17 cues. *Journal of the Acoustical Society of America, 116*, 1351-1354.
18
19 Zeng, F.G., Nie, K., Stickney, G.S., Kong, Y.Y., Vongphoe, M., Bhargave, A., Wei, C., &
20 Cao, K. (2005). Speech recognition with amplitude and frequency modulations.
21 *Proceedings of the National Academy of Sciences USA, 102*, 2293-2298.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60