CrossMark

Applied Psycholinguistics **36** (2015), 463–483
doi:10.1017/S0142716413000313

Changes in voicing perception by adult French speakers after identification training

GREGORY COLLET
Ecole Royale Militaire, Université Libre de Bruxelles, and Fonds de la Recherche Scientifique

CÉCILE COLIN, WILLY SERNICLAES, and INGRID HOONHORST
Université Libre de Bruxelles

EMILY MARKESSIS
Institut Libre Marie Haps

PAUL DELTENRE and JACQUELINE LEYBAERT
Université Libre de Bruxelles

Received: November 23, 2011 Accepted for publication: February 10, 2013

ADDRESS FOR CORRESPONDENCE
Gregory Collet, Unité de recherche en Neurosciences Cognitives and Laboratoire Cognition Langage Développement, Université Libre de Bruxelles, CP 191, 50 Avenue F. D. Roosevelt, Bruxelles 1050, Belgium. E-mail: gcollet@ulb.ac.be

ABSTRACT
The aim of the present study was to investigate changes in voicing identification, discrimination, and categorical perception induced by identification training centered on three different training values. One group of French-speaking adults was trained across a universal auditory boundary (−30 ms voice onset time), and two other groups were trained across arbitrary boundaries (−45 or −60 ms voice onset time). A control group did not receive any training. The results showed that both the −30 and the −45 training groups exhibited a 10 ms shift in the identification boundary. Moreover, for the −30 training group, discrimination and categorical perception changed around the French phonological boundary. These results illustrate the possibility of modifying the French phonological perception after short-time training, particularly when centered on a universal boundary. However, training only had limited effects and even strengthened the phonological boundary, congruent with the hypothesis that this boundary is acquired by a perceptual "coupling" between universal boundaries.

Categorical perception (CP) is a major, although nonspecific, property of speech perception. It is defined as heightened discrimination of an acoustic difference crossing a phonological boundary compared to the same acoustic difference when

© Cambridge University Press 2013 0142-7164/13 $15.00

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

located within a phonological category (Liberman, Harris, Hoffman, & Griffith, 1957). CP is commonly evaluated using identification and discrimination tasks with stimuli that vary along a continuum (for a review, see Damper & Harnard, 2000; Liberman et al., 1957). The relation between these two tasks is evaluated by comparing observed discrimination scores (corresponding to the discrimination responses) to expected discrimination scores (derived from the identification responses). The stronger the relationship between these two scores, the higher the degree of CP. Another categorical property refers to the boundary precision that can be evaluated by measuring either the steepness of the identification curve (Simon & Fourcin, 1978) or the size of the discrimination peak around the phoneme boundary (Wood, 1976). A steeper slope indicates that the boundary, which corresponds to the 50% identification point on the stimulus continuum, is less variable over different trials and can therefore be located with greater precision. Similarly, a higher discrimination peak indicates that the boundary is more frequently located within the same stimulus interval and is therefore more reliable. In addition, the location of the boundary itself is also of interest in indicating the point along the continuum where the change in perception takes place between categories.

Although CP can be observed for all phonological features, voicing perception is the most frequently studied in developmental and training studies. This feature is mainly based on the temporal relationships between two discrete events in stop consonants: the release of the supraglottal closure (burst onset) and the onset of voicing (laryngeal vibrations), that is, voice onset time (VOT; Lisker & Abramson, 1964). Conventionally, negative VOT corresponds to prevoiced stops, whereas positive VOT corresponds to voicing starting after the release.

Studies of human babies younger than 6 months of age (Lasky, Syrdal-Lasky, & Klein, 1975; Streeter, 1976), nonhuman animals (Kuhl & Miller, 1975, 1978), and cross-linguistic perception/production (Lisker & Abramson, 1970) all suggest the existence of two universal boundaries located at −30 and +30 ms VOT (Aslin, Pisoni, Hennessy, & Perey, 1981). However, in Polish, Hebrew, Dutch, French, Spanish, and Arabic (Flege & Eefting, 1986; Horev, Most, & Pratt, 2007; Maassen, Groenen, Crul, Assman-Hulsmans, & Gabreëls, 2001; Serniclaes, 1987; Williams, 1977; Yeni-Komshian, Caramazza, & Preston, 1977), the boundary is located at approximately at 0 ms VOT. This boundary emerges as a conjunction between universal predispositions and exposure to the native language. After 6 months of age, infants raised in French or Spanish environments move from a language-general (with VOT boundaries located at −30 and +30 ms VOT) to a language-specific mode of perception: the 0 ms VOT boundary become relevant in their languages (Spanish: Eilers, Gavin, & Wilson, 1979; Lasky et al., 1975; French: Hoonhorst, Colin, et al., 2009). This finding is congruent with the hypothesis that the 0 ms VOT boundary is acquired by a perceptual "coupling" between the −30 and the +30 ms VOT boundaries, that is, by introducing cross-dependencies between the two boundary locations (Serniclaes, 2011). In this context, coupling means that detection of one feature (e.g., negative VOT) affects the detection of another feature (e.g., changes the positive VOT boundary) such that the boundaries relative to each feature are shifted from the psychoacoustic

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

values (−30 and +30 ms) toward the phonological value (0 ms). This mechanism is similar to *percept–percept couplings* evidenced in visual perception (Epstein, 1982). In some other languages, the phonological boundary straddles one of the two universal boundaries at +30 ms VOT (e.g., for English, see Lisker & Abramson, 1967) or both universal boundaries at +30 and −30 ms VOT (e.g., for Thai, see Abramson & Lisker, 1970; Donald, 1978). These universal boundaries seem to be psychoacoustic in nature and do not always correspond to phonological boundaries in specific languages.

Early studies show that this language-specific mode of perception should be fixed in adulthood and resistant to any changes. This hypothesis was linked to the popular and influential notion that puberty and associated neurological changes represent a cutoff point for second language learning (e.g., De Keyser, 2000; Lenneberg, 1967). It was also supported by the failure in early studies that tried to change adults' phoneme perception through laboratory training (Strange, 1972; Strange & Dittmann, 1984; Vinegrad, 1972). However, over the years, this hypothesis was overturned by a growing body of studies that demonstrated an improvement in discrimination or identification of a specific nonnative contrast after a short period of laboratory training, even in adulthood (Golestani & Zatorre, 2004; Kraus et al., 1995; Tremblay, Kraus, & McGee, 1998; Tremblay, Kraus, McGee, Ponton, & Otis, 2001). Such improvement in performances was also observed in studies addressing the impact of training procedures on nonnative boundaries. For instance, Pisoni, Aslin, Perey, and Hennesy (1982) trained participants to learn a new nonnative category boundary located around −30 ms VOT. They reported that after a 2-hr laboratory training, monolingual speakers of American English were able to reliably label and discriminate voiceless aspirated, voiceless unaspirated, and voiced stops differing in VOT. Hence, they succeeded in altering the perception of voicing of labial stop consonants from a two-way contrast to a three-way contrast. These findings were interpreted as supporting the view that the perceptual mechanisms used by adults in categorizing speech sounds can be easily modified with laboratory techniques in a short period of time. These findings also suggest that the human perceptual system remains at least somewhat malleable over the life span. In a similar vein, Holt, Lotto, and Diehl (2004) trained English-speaking adults to categorize the distribution of tone onset time (a nonspeech analogue of VOT). Groups of participants learned categories in which auditory thresholds were either consistent or inconsistent with the distributional boundaries. Although results showed that participants were able to build categories along the nonspeech continuum, they completed them more quickly in the consistent condition, suggesting that auditory discontinuities at the distributional boundary interact with category learning. These two studies illustrate that it is possible to learn a new nonnative category and that auditory thresholds play an important role in category formation.

Pisoni et al. (1982) did not examine possible perceptual changes when a two-way contrast enters into conflict with a preexisting two-way phonological contrast perception, and Holt et al. (2004) did not investigate the interaction between universal boundary and categorization learning in a preexisting phonological system. However, the present study aims to integrate these two limitations by assessing the evolution of voicing identification, discrimination, and CP after an identification

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

training procedure centered at three different values along the VOT continuum in monolingual French-speaking adults. Our training aimed to introduce a conflict between what participants actually perceive and what training teaches them across three different values within a phonological category. To do so, we observed three different training values along the VOT continuum: one centered on a universal boundary (−30 ms VOT) and two others centered on arbitrary values, one being closer to the phonological boundary (−45 ms VOT) and another further away (−60 ms VOT).

Several reasons have been given in the literature to train participants on the negative side of the continuum. Behavioral (Aslin et al., 1981; Serniclaes, Van Heghe, Mousty, Carré, & Sprenger-Charolles, 2004) and electrophysiological (Hoonhorst, Serniclaes, et al., 2009) studies found an asymmetry in perception between the universal boundary −30 and +30 ms VOT. Several explanations were invoked in the literature. For example, Stevens and Klatt (1974) demonstrated that the underlying reason is the presence of supplementary acoustic cues and, in particular, first formant transitions in positive VOT values, while Burnham, Earnshaw, and Clark (1991) stated that this effect could be explained by a forward masking effect of the high-frequency component (the closure release) by the low-frequency component (the vibration of the vocal cords). To avoid this confound effect, we decided to focus our training along the negative side of the continuum.

For the identification task, we would predict a shift of the phonological boundary initially focused at approximately 0 ms VOT toward the negative side of the continuum and centered on the specific training value after the training sessions. For the discrimination task, we would expect an increase in discrimination across the training value after training; for example, after the training focused on −30 ms VOT, discrimination across −30 ms VOT along the continuum would increase. This expectation is grounded on previous findings by Pisoni et al. (1982) showing that English speakers become sensitive to this new nonnative boundary after training. The same reasoning can be implemented for the −45 and the −60 training groups. However, it was predicted that the discrimination peak across the phonological boundary (0 ms VOT) would be maintained after training and induce a decrease in CP in all training groups. This was expected because the 0 ms VOT French boundary is based on a coupling between the −30 and the +30 VOT universal boundaries. Training French speakers to use the −30 ms VOT would thus also reinforce their 0 ms VOT boundary.

In the present study, all stimuli involved in the training were associated with the same phonological category. Through training, participants had to associate portions of these stimuli with the neighboring category, introducing a mismatch in their perception between what participants actually perceived as /də/ or /tə/ and what training taught them. Using CP (comparing observed and expected discrimination scores) and boundary location analysis, we were able to determine the evolution of different parameters in voicing perception after training. Moreover, with three different trainings, we were able to distinguish training across different values within one specific category and to make a comparison between a universal boundary and an arbitrary boundary.

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

METHOD

Participants

All participants (8 men, 31 women) were native, monolingual French-speaking adults. Ten participants (3 men, 7 women; 18–29 years old, mean = 21.8) were submitted to training centered on the −30 ms VOT value (−30 training group); 10 others (3 men, 7 women; 18–27 years old, mean = 21.1) were submitted to training centered on the −45 ms VOT value (−45 training group); 9 others (2 men, 7 women; 19–27 years old, mean = 22.6) were submitted to training centered on the −60 ms VOT value (−60 training group). The remaining 10 participants (10 women; 18–23 years old, mean = 20) did not receive any training (control group).

Stimuli preparation

The stimuli were alveolar stop + neutral vowel syllables generated by a parallel formant synthesizer (Klatt, 1980) provided by Carré (2004) forming a /də/–/tə/ VOT continuum, from −90 to +90 ms VOT in 15 ms steps. The onset frequencies of the F1, F2, and F3 transitions were 200, 2200, and 3100 Hz, respectively, and the steady-state formant frequencies were 500, 1500, and 2500 Hz, respectively. The F0 value was constant at 120 Hz, and the syllable duration was held constant at 200 ms. This fixed duration of all stimuli implies that when VOT increases, vowel duration decreases. For a stimulus of 90 ms VOT, the vowel duration was 110 ms; for a stimulus of 15 ms VOT, the vowel duration was 185 ms (see Figure 1). This approach with fixed stimulus duration avoids the participant having to answer on the basis of the perception of a difference in duration rather that on the perception of the stimulus as a whole.

The /də/–/tə/ continuum was chosen because the mean VOT boundary is located at approximately 0 ms VOT in French in this context (Hoonhorst et al., 2010; Medina, Hoonhorst, Bogliotti, & Serniclaes, 2010). Previous research conducted by Hoonhorst et al. (2010) using the same continuum showed that French-speaking adults perceived negative VOT stimuli as /də/ and positive VOT values as /tə/.

Procedure

CP was evaluated through identification and discrimination tasks that were carried out before and after training. For the pretraining sessions, half of the participants in each group began with the identification task and the other half with the discrimination task. For posttraining sessions, participants who began with the identification task in the pretraining session began with the discrimination task, and vice versa.

All tasks (identification, discrimination, and training) were programmed and presented using E-Prime (1.1–1.1.4.1) software. The stimuli were binaurally delivered through headphones (Sennheiser HD 202). The participants were seated 50 cm from a 15-in. computer monitor, and all stimuli were presented at 68 dB SPL.

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

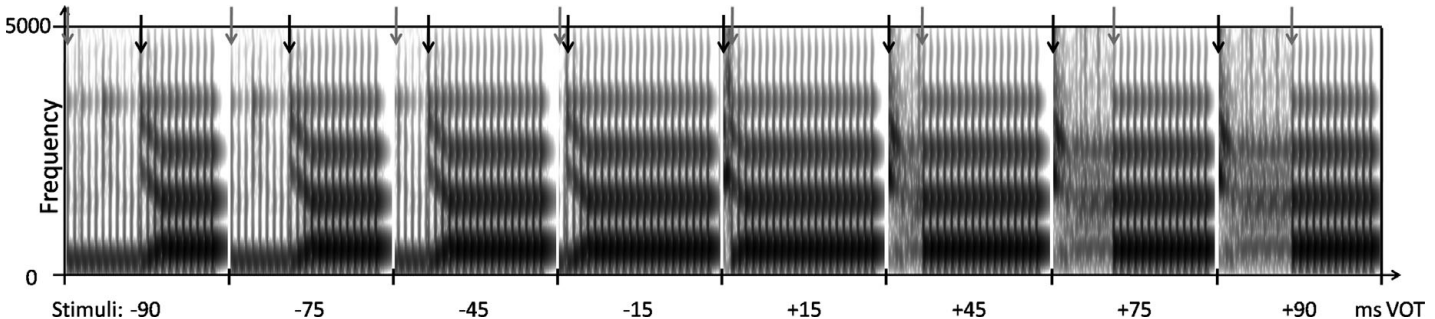


Figure 1. A /də/–/tə/ voice onset time (VOT) continuum spectrogram. The synthetic stimuli varied along a VOT continuum from −90 to +90 ms (from left to right: −90, −75, −45, −15, +15, +45, +75, and +90 ms VOT). The black arrow represents the burst release for each VOT value and the grey arrow the end of the vibration of the vocal cord in the positive VOT and the beginning of the vibration of the vocal cord in the negative VOT.

A familiarization task was presented before each identification and discrimination task to teach participants how to manage the computer and the experiment, and to press on the appropriate key on the keyboard to give their answer.

Pre-/posttesting: Identification task. In the identification task, participants were presented with single tokens chosen randomly from the VOT continuum. For each token, they had to decide whether they heard a /də/ or a /tə/ by pressing the appropriate key on the keyboard. Surrounding the key 1 of the computer, a small card indicated that the key was dedicated to the answer *de* and over the key 0 it was indicated *te*. After each answer, there was a 2-s delay before the next trial was played.

After a familiarization task of 10 random repetitions of each extreme value of the continuum (−90 and +90 ms VOT), a set of 80 stimuli was presented to each participant (10 repetitions of each of the following stimuli: −60, −45, −30, −15, +15, +30, +45, and +60 ms VOT, all presented in a random orders). For the −60 training group, participants were presented with a set of 100 stimuli (10 repetitions of the −75 and +75 ms VOT were added to the previous ones to cover the larger range of VOT involved in the −60 training group). The endpoint stimuli (−90 and +90 ms VOT) were excluded from the experimental stimuli set to avoid any facilitating effect induced by the familiarization task on these stimuli.

Pre-/posttesting: Discrimination task. In the AX discrimination task (Strange, 1992), participants had to decide (by pressing the appropriate key) whether two stimuli of a pair were the same or different. The two stimuli for each pair always differed by 30 ms VOT, and the interstimuli interval (ISI) was 100 ms. A short ISI was used in order to trigger direct comparisons in the discrimination task based on acoustic, rather than phonological, differences between stimuli that suggest the activation of low-level auditory processes, which normally occur when the ISI is less than 500 ms (Werker & Logan, 1985). Shorter ISIs would trigger direct comparisons based on acoustic (rather than phonological) differences between the stimuli, which would suggest the activation of low-level auditory processes.

A 2-s intertrial interval separated a given trial from the previous one. The VOT values were chosen to test the sensitivity to stimuli pairs straddling each of the following potential perceptual boundaries: 0 ms VOT (phonological boundary); −30 and +30 ms VOT (universal boundaries); and −45, +45, −60, and +60 ms VOT (arbitrary values).

A familiarization task was presented to each participant using the extreme values of the continuum (−90 and +90 ms VOT). A single set of 20 pairs (five presentations of each of the following pairs: −90/−90, −90/+90, +90/−90, and +90/+90 ms VOT) was presented to each participant in a different random order. Each participant was then exposed to a set of 90 pairs presented in a different random order (five repetitions of each of the eight identical pairs: −60/−60, −30/−30, −45/−45, −15/−15, +15/+15, +45/+45, +30/+30, and +60/+60 ms VOT and five repetitions of each of the 10 different pairs: −60/−30, −30/−60, −45/−15, −15/−45, −15/+15, +15/−15, +15/+45, +45/+15, +30/+60, and +60/+30 ms VOT). For the −60 training group, five repetitions of each of the

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

following pairs were added: −75/−75, −45/−75, −75/−45, +45/+75, +75/+45, and +75/+75 ms VOT (for similar reasons as used in the identification task).

Training. The participants were trained using an identification task called "perceptual fading" (Jamieson & Morosan, 1986). In this training method, we progressively reduced the acoustic distance (magnitude of VOT difference in a pair) between two stimuli as a function of performance, which allows participants who successfully learn to distinguish the endpoint stimuli to attempt to distinguish contrasts with smaller acoustic step size and, thus, to learn more subtle perceptual aspects of the acoustic parameters. Over the course of 2 weeks, participants underwent five 1-hr training sessions with feedback provided every 2 days.

During each session and for each training group (−30, −45, or −60), participants had to perform the largest number of blocks of 20 stimuli in 1 hr. In each block, the 2 stimuli were presented 10 times in random order and could be achieved in at least 2 min, depending on the speed of each participant. It was therefore possible to achieve a maximum of 30 blocks in one session.

Each participant always began the session at Level 1 (Figure 2) with a 50 ms VOT difference between the two stimuli and centered on the trained value (−30, −45, or −60 ms VOT). When the participants achieved the criterion, which was defined as a minimum of 18/20 correct responses for three consecutive blocks, the acoustic distance was reduced to 40 ms (Level 2), then to 30 ms (Level 3), then to 20 ms (Level 4), and finally to 10 ms (Level 5: see Figure 2). When participants were able to reach this criterion for all five levels, the training session ended, even though the 1-hr training session was not over.

The participants' task was to decide whether the stimulus was a /də/ or a /tə/ by pressing the appropriate response key. For the −30 training group, /də/ was considered the correct answer for all stimuli with VOT values below −30 ms VOT, and /tə/ was considered the correct answer for all stimuli with VOT values above −30 ms VOT. In this way, we introduced a mismatch with French phonological perception in which stimuli between −30 and 0 ms VOT are perceived as /də/. The same principle was applied to the −45 and the −60 training groups.

Visual feedback was presented on the computer screen after each trial to indicate whether the response was correct (green screen) or incorrect (red screen). Green feedback appeared for the −30 training group when the subject correctly identified the −25, −20, −15, −10, and −5 ms VOT stimuli as /tə/ and the −35, −40, −45, −50, and −55 ms VOT stimuli as /də/. For the −45 training group, a green reinforcement screen appeared when the subject correctly identified the −40, −35, −30, −25, and −20 ms VOT stimuli as /tə/ and the −50, −55, −60, −65, and −70 ms VOT stimuli as /də/. For the −60 training group, a green reinforcement screen appeared when the subject correctly identified the −55, −50, −45, −40, and −35 ms VOT stimuli as /tə/ and the −65, −70, −75, −80, and −85 ms VOT stimuli as /də/.

Data processing and statistical analysis

The slope of the identification functions was assessed separately for each participant using logistic regression (McCullagh & Nelder, 1983) with the labeling

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

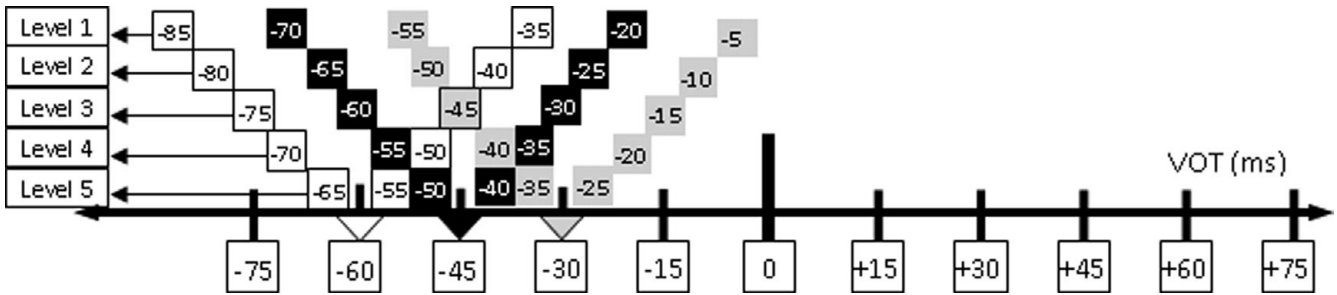


Figure 2. The voice onset time (VOT) values used in the experiment. The white squares refer to the stimuli used in the pre- and posttraining identification and discrimination task. For the training sessions, the values centered on −30, −45, and −60 ms VOT are represented by the grey, black, and white squares, respectively. For all training types, the first level is associated with the largest acoustic distance between the two stimuli in the pair. As the level increases from 1 to 5, the acoustic distance decreases (from 50 to 10 ms VOT).

response as the dependent variable and VOT stimulus as the independent variable. The logistic function (McCullagh & Nelder, 1983) has frequently been used for fitting identification curves in studies of speech perception (Nearey, 1990). Boundary location was then computed separately for each participant from the intercept and slope values (for extended information, see Hoonhorst et al., 2010; Medina et al., 2010).

The identification results were tested with two repeated measures analyses of variance (ANOVAs) with Session as the within-subject factor (two levels: pre- and posttraining) and Group as the between-subject factor, with either the slope or the boundary of the identification curve as the dependent variable. Corrections for multiple tests were not performed because comparisons between groups were planned.

CP was assessed by comparing the observed discrimination d' scores with those from the expected discrimination derived from the identification data computed with elementary probability formulas (Pollack & Pisoni, 1971; for extended information, see Hoonhorst et al., 2010; Medina et al., 2010). Observed and expected discrimination were transformed into d' by summing the normal deviate z values corresponding to the score of "same" responses to the "same" pairs and those corresponding to the score of "different" responses to the "different" pairs. To circumvent the problem that z (100%) and z (0%) equal infinity, scores above 50% were reduced and scores below 50% were increased by 2.5%, following the procedure described by Macmillan and Creelman (2005).

The observed discrimination results were tested with a repeated measures ANOVA with Session (two levels: pre- and posttraining) and VOT (five levels: −45, −30, 0, +30, and +45 VOT) as within-subject factors and Group as the between-subject factor, with d' scores as the dependent variable. The relation between the expected and observed discrimination results for each training group was tested in a repeated measures ANOVA with Task (two levels: expected and observed discrimination), Session (two levels: pre- and posttraining), and VOT (five levels: −45, −30, 0, +30, and +45 VOT) as within-subject factors and Group as the between-subject factor, with d' scores as the dependent variable.

Whenever the Mauchly sphericity test was significant, Greenhouse–Geisser corrected degrees of freedom values were used.

RESULTS

Identification

The mean identification curves for each of the four groups are presented in Figure 3. For the mean boundary, significant effects of session, $F(1, 3) = 26.9$; $p < .001$; $\eta^2 = 0.43$, and group, $F(3, 35) = 4.6$; $p = .008$; $\eta^2 = 0.28$, and a significant Session × Group interaction, $F(3, 35) = 4.6$; $p = .008$; $\eta^2 = 0.28$, were observed. For each group, the mean boundary score in pretraining was compared to that in posttraining. The analysis revealed a significant session effect for the −30 training group, $F(1, 9) = 29.9$; $p < .001$; $\eta^2 = 0.77$, and the −45 training group, $F(1, 9) = 14.1$; $p = .004$; $\eta^2 = 0.6$. No effect was reported for either the −60 training group

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

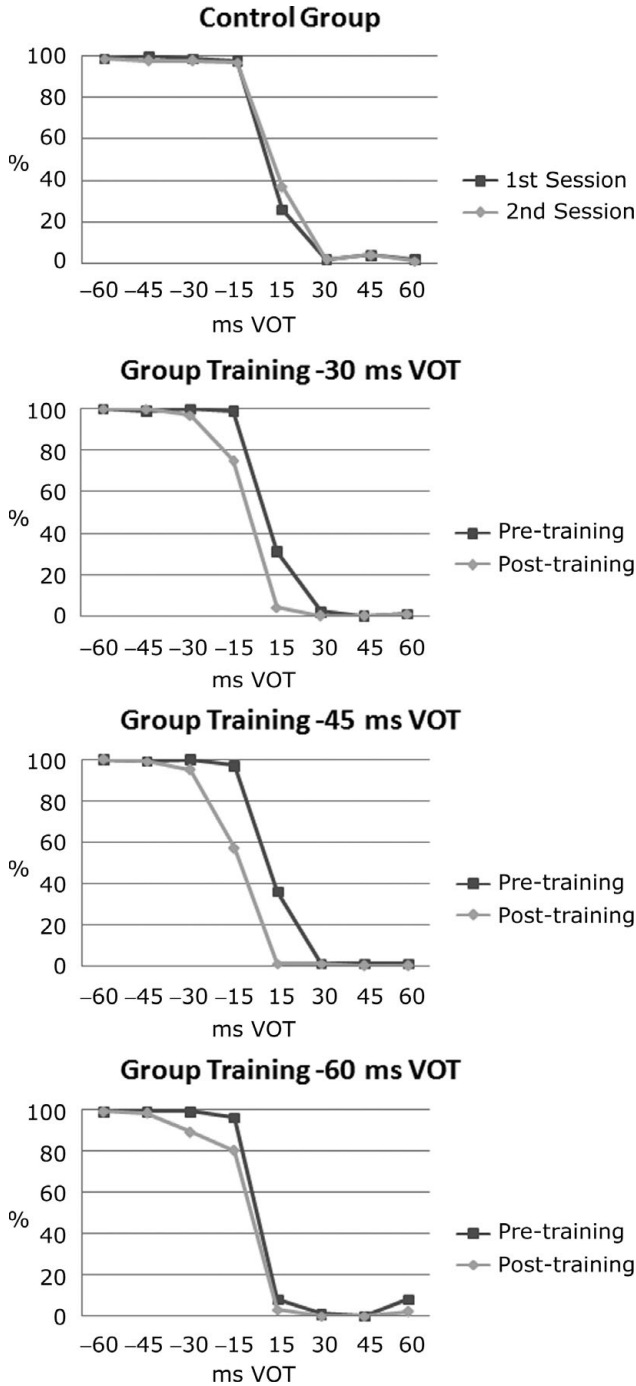


Figure 3. The identification function curve on the /də–tə/ voice onset time (VOT) continuum in (dark grey) pre- and (light grey) posttraining for the four groups: control group, and the −30,−45, and −60 training groups. The curve corresponds to identification of /də/ (%).

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

Table 1. *The boundary location values reported for each training group in pretraining and posttraining*

	Pretraining	Posttraining
−30 Training group	12	3
−45 Training group	12	1
−60 Training group	9	3
Control group	10.5	12

Note: All values are expressed in milliseconds of voice onset time (ms VOT).

or the control group, $F\,(1, 9) = 3.1$; $p = .11$; $\eta^2 = 0.3$ and $F < 1$, respectively. All boundary values in pretraining and posttraining are reported in Table 1.

Discrimination

Observed discrimination responses. Examination of Figure 4 reveals an increase in the 0 ms VOT discrimination peak between the pre- and posttraining session for the −30 training group, whereas no such increase was present between the first and second sessions for the control group.

In the ANOVA comparing the −30 training group and the control group, the Session × VOT × Group interaction was significant, after Greenhouse–Geisser correction: $F\,(4, 72) = 4.29$; $p = .004$; $\eta^2 = 0.19$. The "across versus within-category" planned interaction contrast, that is, the between-group difference in the change from the pre- to posttraining value of the 0 ms discrimination peak, was also significant, $F\,(1, 18) = 7.48$; $p = .02$; $\eta^2 = 0.29$. When tested separately for each group, the Session × VOT interaction was significant for the −30 training group, after Greenhouse–Geisser correction: $F\,(4, 36) = 5.26$; $p = .002$; $\eta^2 = 0.37$, but was not significant for the control group ($F < 1$). The "across versus within-category" planned interaction contrast, that is, the change from the pre- to posttraining value of the 0 ms discrimination peak, was significant, $F\,(1, 9) = 6.86$; $p = .03$; $\eta^2 = 0.43$, for the −30 training group. There were no important changes in the magnitude of the 0 ms VOT discrimination peak between the pre- and posttraining sessions for either the −45 or the −60 training groups. The Session × VOT × Group interaction was not significant for the comparisons between either the −45 training group or the −60 training group and the control group, $F\,(4, 72) = 1.35$; $p = .26$; $\eta^2 = 0.07$; $F < 1$, respectively.

Expected versus observed discrimination responses. Figure 4 shows that for all four groups, CP was far from perfect during the first session: the observed discrimination scores were larger than the expected scores for the + 30 ms VOT within-category pair for all four groups, and the observed discrimination was lower than expected for the 0 ms VOT between-category pair for the −30 and the −60 training groups. Concerning the effect of training on CP, Figure 4 shows that for the -30 training group, there was a decrease in CP after training owing to

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

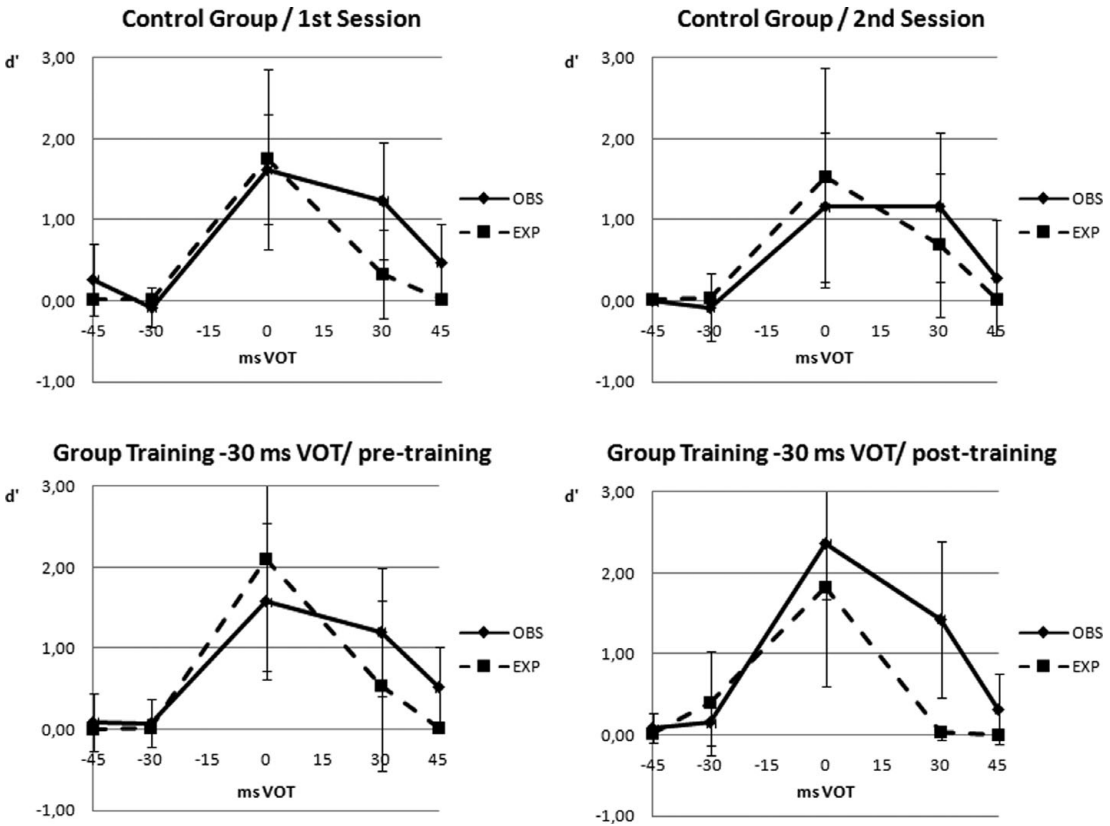


Figure 4. The expected (EXP) and observed (OBS) *d′* scores on the /də–tə/ voice onset time (VOT) continuum for the control group and the −30, −45, and −60 training groups. Pretraining scores are presented on the left side and posttraining scores on the right side. Error bars represent the standard deviation. The *x* axis values are the VOT values (ms).

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

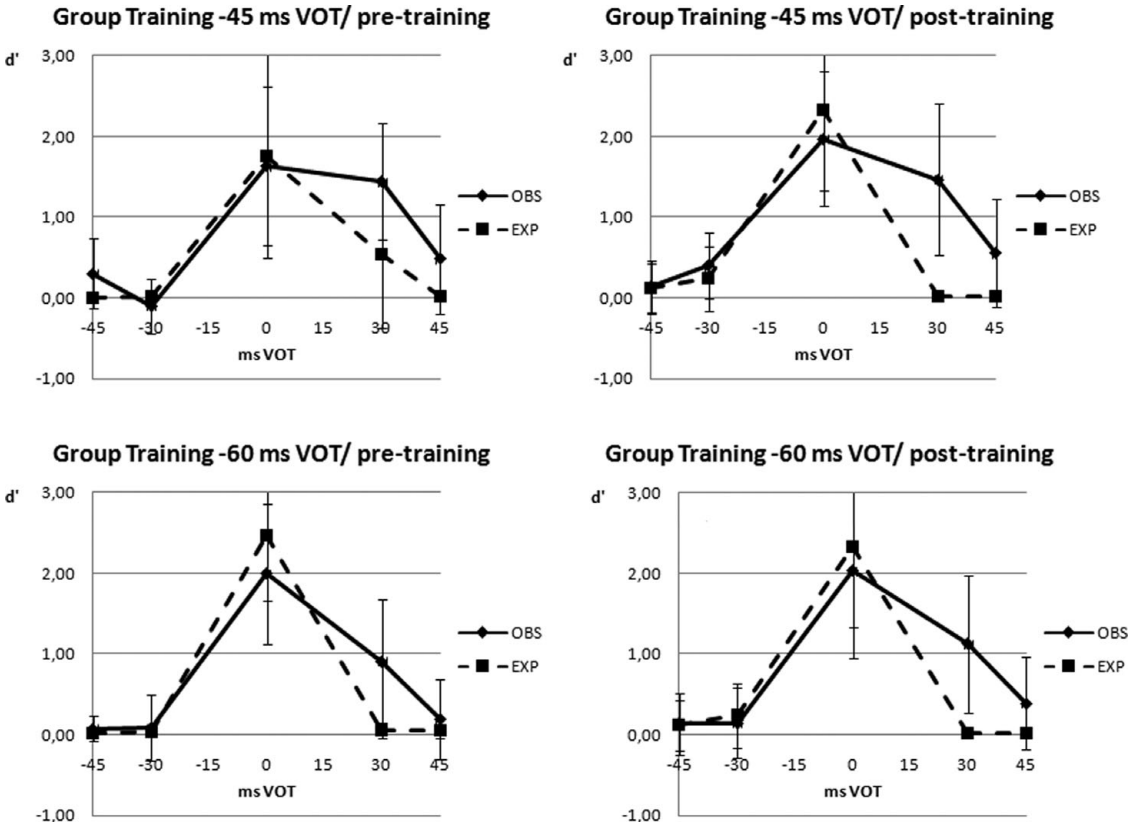


Figure 4 (*cont.*)

larger differences between observed and expected discrimination at +30 ms VOT. By contrast, there was no substantial change in CP between the first and second sessions for the control group; the difference between observed and expected discrimination increased slightly for the 0 ms VOT pair but decreased for the +30 ms VOT pair.

The decrease in CP with the −30 training group was confirmed by the results of a repeated measure ANOVA. The Task × Session × VOT × Group interaction was significant after Greenhouse–Geisser correction: $F\ (4, 72) = 3.62$; $p = .01$; $\eta^2 = 0.17$. However, the "across versus within-category" planned interaction contrast was not significant, $F\ (1, 18) = 2.24$; $p = .15$; $\eta^2 = 0.11$, indicating that the decrease in CP was due to both across- and within-category changes, that is, to larger differences between the observed and expected discrimination scores at both 0 and +30 ms VOT. When tested separately for each group, the Session × Task × VOT interaction was just below significance for the −30 training group, after Greenhouse–Geisser correction: $F\ (4, 36) = 3.74$; $p = .06$; $\eta^2 = 0.29$, and the "across- versus within-category" planned interaction contrast, that is, the between-group difference in the change from the pre- to the posttraining value of the difference between the expected and the observed 0 ms discrimination peaks, was not significant, $F\ (1, 9) = 2.70$; $p = .13$; $\eta^2 = 0.23$.

Concerning the −45 and the −60 training groups, examination of Figure 4 suggests that there were no substantial differences in CP between the pre- and the posttraining sessions. This finding is confirmed by the results of repeated measures ANOVAs showing that the Task × Session × VOT × Group interaction was not significant for comparisons between either the −45 training group or the −60 training group and the control group (both $F < 1$).

Training

At the end of training, all participants from all groups reached the criterion for the three first levels. For the −30 training group, seven participants were able to reach the fourth level, and only three were able to reach the fifth level. For the −45 training group, four participants were able to reach the fourth level, and six were able to reach the fifth level. For the −60 training group, four participants were able to reach the third level, one reached the fourth level, and four were able to reach the fifth level.

DISCUSSION

The major aim of this study was to assess the effect of identification training on voicing identification, discrimination, and CP in monolingual French-speaking adults. This identification training aimed to introduce a conflict between what participants actually perceive and what training teaches them across three different values within a phonological category. We observed the evolution of three different training values along the VOT continuum: one centered on a universal auditory boundary (–30 ms VOT) and two others centered on arbitrary values, one fairly close to the phonological boundary (–45 ms VOT) and another further away (–60 ms VOT).

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

Identification: Boundary location

The results of the identification task showed that, after training, the phonological boundary shifted from approximately 10 ms VOT for the −30 and −45 training groups but did not shift for the −60 training group and the control group. These results suggest that the 5 hr of perceptual fading training were sufficient to modify the value of the phonological voicing boundary of adult French speakers. Such modification for the −30 and the −45 training groups could be understood by the training itself. Both the −30 and the −45 training groups exhibited shifts in the identification boundary in the expected direction. No effects on the slope of the identification curves were observed, indicating that the boundary shifts did not affect the *precision* of the identification boundary. The new boundary was, thus, no less precise in any training groups than was the original, which is a good indication of the quality of the training. However, the absence of shift for the −60 training group is more difficult to explain. This noneffect could be explained by numerous factors. However, if we look carefully at the different VOT values included in the training, we may notice this absence any time the −30 ms VOT value is involved. As we will discuss in the next section, we may suspect that involvement of a universal boundary at any point in the training may induce learning effects.

The long-term effects of this training will certainly remain tenuous. Once the training ends, the participant is again immersed in an environment where no conflict exists in his phonological perception. In this case, the 10 ms VOT shift should quickly fade, and the phonological boundary should resume its initial value. Unfortunately, the long-term effects were not investigated in the present study but will be important to include in future studies.

Discrimination and CP

The results of the discrimination task showed a significant increase (10.5%) in the number of correct discriminations across the French phonological boundary (0 ms VOT) for the −30 training group. Moreover, the CP (correspondence between expected and observed discrimination responses) showed a significant decrease only for the −30 training group, owing to both across- and within-category changes (at both 0 and +30 ms VOT). These results suggest that the 5 hr of perceptual fading training are sufficient to modify discrimination and CP of adult French speakers and that the conflict introduced in perception across the universal boundary had a large impact on the acoustic space compared to that of the other trained values along the VOT continuum. However, we should keep in mind that these training effects will probably vanish through exposure to environmental language input. An open question is to see whether training effects might facilitate the acquisition of the phonological boundaries prevailing in another language during immersion in this language.

As mentioned in the introductory section, the −30 ms VOT is one of two universal auditory thresholds. Children under 6 months of age raised in a Spanish- (Lasky et al., 1975), English- (Aslin et al., 1981; Eimas, 1975), or French-speaking environment (Hoonhorst, Colin, et al., 2009) reacted to both −20/−40 ms VOT and +20/+40 ms VOT contrasts regardless of whether these contrasts straddled the

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

phonological boundary used in their language environment. Moreover, sensitivity to the −30 and +30 ms boundaries has been found in animal studies (Sinnott & Gillmore, 2004), psychoacoustic studies (Pisoni, 1977), and neurophysiological investigations of humans (Hoonhorst, Serniclaes, et al., 2009).

Even though the −30 and the −45 training groups exhibited shifts in the identification boundary in the expected direction, significant training effects on discrimination were only obtained for the −30 training group, suggesting that the training for this boundary had deeper effects than that for −45 ms VOT. This difference in shift was expected from the difference in the auditory status between these boundaries, because −30 ms VOT corresponds to a universal boundary, whereas −45 ms VOT has no special value. Overall, it appears that the closer the training value is to −30 ms VOT, the stronger the effects are according to different criteria (identification shift and change in discrimination scores). However, note that the training does not improve identification and discrimination around the −30 ms boundary. Rather, the training only displaces the identification boundary and discrimination peak toward this value, which suggests that the phonological boundary is strongly anchored around its original value (0 ms VOT) and that possible training effects are constrained by the universal boundaries. This phenomenon can be understood if it is assumed that the French phonological boundary (at 0 ms VOT) arises from a coupling of the universal boundaries (at −30 and +30 ms VOT; Hoonhorst, Colin, et al., 2009; Serniclaes, 2011).

These results support previous studies showing that categorical boundaries can be acquired (Holt et al., 2004) or modified (Pisoni et al., 1982) following a training procedure within limits imposed by auditory thresholds. However, this study specifies that a phonological boundary is sensitive not only to conflict introduced across auditory thresholds but also to certain arbitrary values near the phonological boundary. Auditory thresholds interact with category shifting but not in a specific way, given that shifts occurred for the group trained around the −30 ms VOT auditory threshold and for the group trained around the −45 ms arbitrary VOT value. However, for discrimination and CP (assessed through comparison between identification and discrimination tasks), this interaction was specific, with changes occurring only for the −30 training group. These results clearly support that modifications in perception (CP and boundary location) can occur even in adulthood.

The present study outlined two methodological novelties. The first showed the importance of using identification and discrimination tasks in order to take into account the relation between these two tasks (CP) for assessing the effects of the training, even if the training was only based on an identification task. Identification training affected both the identification and the discrimination performances, although differently. This finding is congruent with the assumption that these two tasks involve distinct mechanisms that do not reflect the same level of speech representation (Wood, 1976). The second methodological novelty concerned the range of stimulus values taken into account for assessing the training effects. Up until now, most studies (Golestani & Zatorre, 2004; Kraus et al., 1995; Tremblay et al., 1998, 2001) have evaluated the effects of training only on the trained contrasts. In the present study, even if the training was centered on three specific values, we measured training effects on the CP along the entire voicing continuum

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

by comparing observed discrimination scores to expected discrimination scores. We showed that the training effects were not restricted to the trained VOT values, illustrating the importance of regarding the phenomenon of learning as a general mechanism that modifies a large part of the perceptual space near the listener's native-language phonological boundary. This phenomenon holds especially true for studies that evaluate these capacities in children with specific language impairments or dyslexia who might suffer from CP troubles (Manis & Keating, 2004; Serniclaes et al., 2004; Serniclaes, Sprenger-Charolles, Carré, & Démonet, 2001). Specifically, the CP deficit in dyslexia arises from an increased sensitivity to universal auditory boundaries that are not phonological in their language but may be in other languages, a phenomenon called "allophonic perception" (for voicing: Bogliotti, Serniclaes, Messaoud-Galusi, & Sprenger-Charolles, 2008; for place of articulation: Dufor, Serniclaes, Sprenger-Charolles, & Démonet, 2009; Noordenbos, Segers, Serniclaes, Mitterer, & Verhoeven, 2012). Hence, to improve CP in children with language problems, an appropriate training procedure, centered on the phonological boundary, must be selected, which has already been tested with some success (Collet et al., 2012).

This study showed that CP and phonological boundaries can be modified using an appropriate training procedure despite limitations depending on the training value. Because this training is efficient for adult speakers, it seems promising for remediating phonological deficits in language-impaired children (i.e., those with specific language impairment or dyslexia).

ACKNOWLEDGMENTS

This work was financially supported by FER Grant EC9014000007 from the Université Libre de Bruxelles (to C.C.); by FRFC Grant 2.4621.07 (to J.L.); by an FER grant from the Université Libre de Bruxelles (Faculty of Medicine), the Brugmann Foundation, the David and Alice Van Buuren foundation (to P.D.); by the Gosset Foundation (to E.M.); and by ANR Grant ANR-07-BLAN-0014-01 (France; to W.S.). The authors are grateful to René Carré (CNRS Researcher at ENST Paris) for providing the speech synthesis software.

REFERENCES

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voice onset time continuum: Cross-language tests. In B. Hala, M. Romporti, & P. Jonota (Eds.), *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569–573). Ismaning, Germany: Hueber.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. V. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development, 52*, 1135–1145.

Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., & Sprenger-Charolles, L. (2008). Discrimination of speech sounds by dyslexic children: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology, 101*, 137–175.

Burnham, D. K., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: Infants children and adults. *Journal of Child Language, 18*, 231–260.

Carré, R. (2004). Program SyntFormVoy [Computer software]. Lyon, France: Laboratoire Dynamique du Language, CNRS.

Collet, G., Colin, C., Serniclaes, W., Hoonhorst, I., Markessis, E., Deltenre, P., et al. (2012). Effect of phonological training in French children with SLI: Perspectives on voicing identification,

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

discrimination and categorical perception. *Research in Developmental Disabilities, 33*, 1805–1818.
Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception & Psychophysics, 62*, 843–867.
DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499–534.
Donald, S. L. (1978). *The perception of voicing contrasts in Thai and English*. Unpublished doctoral dissertation, University of Connecticut.
Dufor, O., Serniclaes, W., Sprenger-Charolles, L., & Démonet, J.-F. (2009). Left pre-motor cortex and allophonic speech perception in dyslexia: A PET study. *NeuroImage, 46*, 241–248.
Eilers, R., Gavin, W., & Wilson, W. (1979). Linguistic experience and phonetic perception in infancy: A cross-linguistic study. *Child Development, 50*, 14–18.
Eimas, P. D. (1975). Speech perception in early infancy. In L. Cohen & P. Salapatek (Eds.), *Infant perception: Vol. 2. From sensation to cognition* (pp. 193–231). New York: Academic Press.
Epstein, W. (1982). Percept-percept coupling. *Perception, 11*, 75–83.
Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica, 43*, 155–171.
Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage, 21*, 494–506.
Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America, 116*, 1763–1773.
Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2009). French native speakers in the making: From language-general to language-specific voicing boundaries. *Journal of Experimental Child Psychology, 104*, 353–366.
Hoonhorst, I., Medina, V., Colin, C., Markessis, E., Radeau, M., Deltenre, P., et al. (2010). The development of categorical perception: Comparisons between voicing, colors and facial expressions. *Speech Communication, 53*, 417–430.
Hoonhorst, I., Serniclaes, W., Collet, G., Colin, C., Markessis, E., Radeau, M., et al. (2009). N1b and Na subcomponents of the N100 long latency auditory evoked-potential: Neurophysiological correlates of voicing in French-speaking subjects. *Clinical Neurophysiology, 120*, 897–903.
Horev, N., Most, T., & Pratt, H. (2007). Categorical perception of speech (VOT) and analogous non speech (FOT) signals: Behavioral and electrophysiological correlates. *Ear and Hearing, 28*, 111–128.
Jamieson, D. G., & Morosan, D. E. (1986). Training nonnative speech contrasts in adults: Acquisition of the English /delta/–/theta/ contrast by francophones. *Perception & Psychophysics, 40*, 205–215.
Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971–995.
Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience, 7*, 25–32.
Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants. *Science, 190*, 69–72.
Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America, 63*, 905–917.
Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environment. *Journal of Experimental Child Psychology, 20*, 215–225.
Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
Liberman, A. M., Harris, K. S., Hoffman, K. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368.
Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.
Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10*, 1–28.

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In B. Hala, M. Romporti, & P. Jonota (Eds.), *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563–567). Ismaning, Germany: Hueber.

Maassen, B., Groenen, P., Crul, T., Assman-Hulsmans, C., & Gabreëls, F. (2001). Identification and discrimination of voicing and place of articulation in developmental dyslexia. *Clinical Linguistics and Phonetics, 15*, 319–339.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. London: Erlbaum.

Manis, F., & Keating, P. (2004). Speech perception in dyslexic children with and without language impairments. *UCLA Working Papers in Phonetics, 103*, 30–47.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London: Chapman & Hall.

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in French: Comparing adults, adolescents and children. *Journal of Phonetics, 38*, 493–503.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics, 18*, 347–373.

Noordenbos, M. W., Segers, E., Serniclaes, W., Mitterer, H., & Verhoeven, L. (2012). Neural evidence of allophonic perception in children at risk for dyslexia. *Neuropsychologia, 50*, 2010–2017.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two components tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61*, 1352–1361.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennesy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology. Human Perception and Performance, 8*, 297–314.

Pollack, S., & Pisoni, D. (1971). On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science, 24*, 299–300.

Serniclaes, W. (1987). *Etude expérimentale de la perception du trait de voisement des occlusives du français* [Experimental study on the perception of the voicing feature in French stop consonants]. Unpublished doctoral dissertation, Université Libre de Bruxelles.

Serniclaes, W. (2011). Features are phonological transforms of natural boundaries. In G. N. Clements & R. Ridouane (Eds.), *Cognitive, physical and developmental bases of distinctive speech categorie*. London: John Benjamins.

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Démonet, J. F. (2001). Perceptual categorization of speech sounds in dyslexics. *Journal of Speech, Language, and Hearing Research, 44*, 384–399.

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology, 87*, 336–361.

Simon, C., & Fourcin, A. J. (1978). Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America, 63*, 925–935.

Sinnott, J. S., & Gilmore, C. S. (2004). Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perception & Psychophysics, 66*, 1341–1350.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America, 55*, 653–659.

Strange, W. (1972). *The effects of training on the perception of synthetic speech sound: Voice onset time*. Unpublished doctoral dissertation, University of Minnesota.

Strange, W. (1992). Learning non-native phoneme contrasts: Interaction among subject, stimulus and task variables. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 197–219). Tokyo: Ohmsha.

Strange, W., & Dittman, S. (1984). Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131–145.

Streeter, L. A. (1976). Language perception of 2 months old infants show effects of both innate mechanisms and experience. *Nature, 259*, 39–41.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time-course of auditory perceptual learning: Neurophysiological changes during speech–sound training. *NeuroReport, 9*, 3556–3560.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1–P2 complex after speech–sound training. *Ear and Hearing, 22*, 79–90.

Vinegard, M. D. (1972). A direct magnitude scaling method to investigate categorical vs. continuous modes of speech perception. *Language and Speech, 15*, 114–121.

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*, 35–44.
Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics, 5*, 169–184.
Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America, 60*, 1381–1389.
Yeni-Komshian, G. H., Caramazza, A., & Preston, M. S. (1977). A study of voicing in Lebanese Arabic. *Journal of Phonetics, 5*, 35–48.

https://doi.org/10.1017/S0142716413000313 Published online by Cambridge University Press