Effects of Categorical Learning on the Auditory Perceptual Space

Doctoral Thesis

by
Valeria C. Caruso
Supervisor
Professor Evan Balaban

Cognitive Neuroscience Sector
SISSA/ISAS
International School for Advanced Studies
Via Stock 2/2, 34135 Trieste, Italy

2009

Acknowledgment

It is a pleasure to thank those who made this thesis possible.

I am heartily grateful to my supervisor, Professor Evan Balaban, for his invaluable guidance, support, encouragement and friendship throughout the project. Without his help, this work would not have been possible.

I would also like to thank the rest of my thesis committee: Professor Piero Paolo Battaglini, Professor Giuseppe Longobardi, Professor Raffaella Rumiati, Professor Michael Siegal.

In the course of my years in graduate school I was in contact with many students, researchers, professors who have contributed towards my understanding and thoughts. In particular, I wish to express my gratefulness to Professor Mathew Diamond who taught me more than he is aware of and to Professor John Nicholls whose wise suggestions were deeply appreciated.

My sincere appreciation and friendship go to the fellow graduate students and postdocs, especially Damir Kovacic, Mahan Azadpour, Linda Sebastianutto, Katya Vinnik, Juan Pedro Vargas, Nicola Van Rijsbergen, Yaron Meirovich, Alessandro Cicerale who enriched my time at SISSA.

I would like to thank the Cognitive Neuroscience Sector of SISSA for providing laboratory facilities. My sincere appreciation is due to SISSA staff, especially to Marijana Sjekloca, Alessio Isaja, Lorena Quadarella and Andrea Sciarrone who have provided assistance in various occasions. Special thanks also to all the participants in the experiments for their time and contribution.

Above all, I owe my deepest gratitude to my precious friends in the Math and Physics Sectors at SISSA, who made me feel at home in Trieste and to my friends and family outside Trieste who were always there for me and to whom I did not dedicate all the time and attention they deserved.

Index

Chapter 1

1.1 Categorical perception

Categorical perception is a psychophysical effect in which the way some physical stimuli are categorized influences the way they are perceived. Specifically, perceptual distances between pairs of stimuli reflect their position relative to some category border rather than their physical distance, so that between-category discriminations are more accurate than equivalent within-category discriminations (Livingston, Andrews & Harnad, 1998).

Categorical perception (CP) was first observed in perceptual experiments with speech sounds (Liberman et al., 1957) and colors (e.g. Bornstein, 1987), but it has since been found in a variety of domains and with stimuli of variable complexity. For example, CP has been found for visual textures (Pevtzow & Harnad, 1997), visual forms (induced by training, Livingstone et al., 1998) and different features of faces, like expression of emotions (e.g. Etcoff et al., 1992), race (e.g. Levin & Angelone, 2002), gender (e.g. Campanella et al., 2001), and individual identity (e.g. Beale & Keil, 1995).

1.1.1 Color perception

In color perception the wavelength of light in the visible spectrum range varies continuously, but is perceived as subdivided into color categories that appear to affect color discrimination: the discrimination of colors separated by a category boundary, for example a shade of blue and a shade of green, is faster and more reliable than the discrimination of equivalently spaced within-category colors, for example two shades of green (e.g. Bornstein & Korda, 1984; Roberson et al., 2000).

Color categorical perception seems to be built-in in the visual perceptual system (Berlin & Kay, 1969, cited in Bornstein 1897; Heider, 1972; Rosch, 1973; Mervis at al., 1975) although the neurological basis of color categorization has not yet been identified (Valberg, 2001) and there is some evidence that training and language might influence color perception (e.g. Roberson et al., 2000, Özgen & Davies, 2002; Pilling et al., 2003).

1.1.2 Speech perception

Categorical perception in speech has been found for several dimensions of phonetic contrast, such as place of articulation, voicing and manner of articulation (e.g. Liberman, Harris, Hoffman & Griffith, 1957; Liberman et al., 1961; Ambramson and Lisker, 1970) and to a lesser extent for vowels (Fry et al., 1962; Stevens, 1969).

The results of the first experiment demonstrating CP for speech sounds are shown in figure 1.1 (Liberman, Harris, Hoffman & Griffith, 1957). With equal acoustic differences, listeners discriminated better between speech sounds which they normally classified as being different phonemes, compared to the sounds which they normally classified as being the same phoneme. The stimuli were two-formant approximations to consonant-vowel syllables, and they varied in the extent and direction of the second formant transitions, which are the important cues for the perception of the voiced stops /b/, /d/, and /g/. Liberman named this effect "categorical perception" (Liberman et al., 1967).

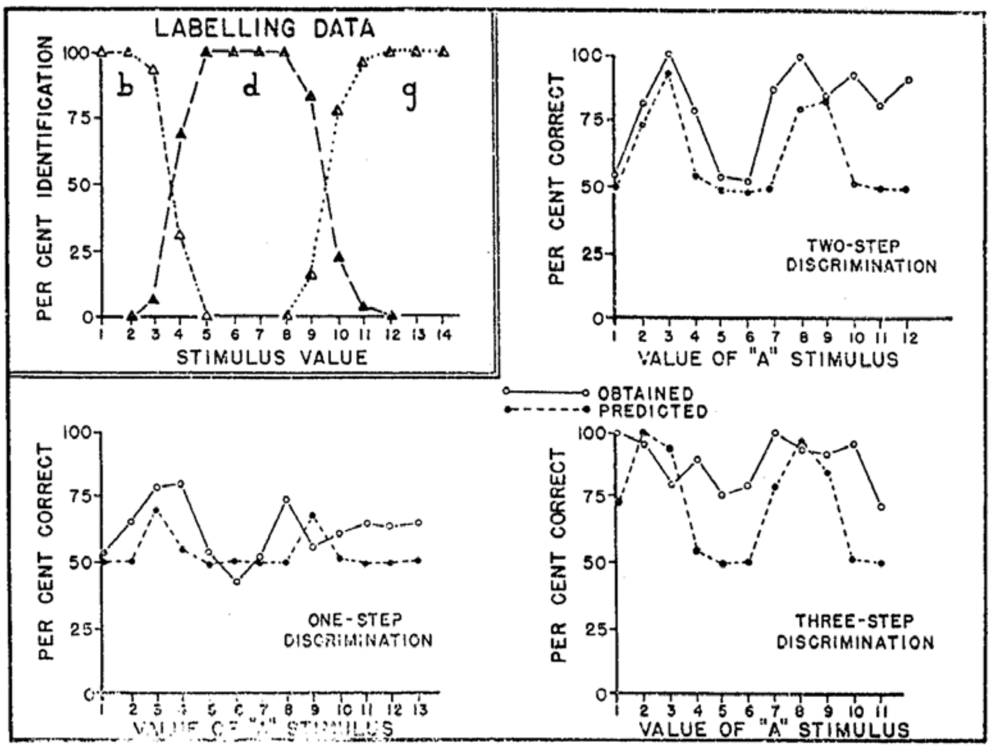


FIGURE 1.1 Identification and discrimination data for a single subject. From Liberman et al. 1957

In the **Motor Theory of Speech Perception**, Liberman and colleagues (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967) suggested that CP is unique to speech perception, and that it is part of a specific mechanism based on motor production and dedicated to processing speech sounds, as opposed to other auditory stimuli.

This interpretation was subsequently challenged by the findings that CP can occur for non-speech sounds, such as musical intervals (Burns and Ward, 1978) and noise onsets (Miller et al., 1976), and that CP was also shown for speech sounds by animals that are unable to produce those sounds, like monkeys, chinchillas and some species of birds (Kuhl & Miller, 1975, 1978; Kuhl, 1981; Kuhl & Padden, 1982, Dooling et al., 1995; see also Diehl et al., 2004 for a review).

Four-month-old infants have been shown to have speech CP before they begin to speak (Eimas et al., 1971; Kuhl, 1993), although learning to speak seems to refine their phonemic categories (Werker & Tees, 1984).

Different mechanisms may underlie CP for different stimuli and in different modalities. In the auditory domain some discontinuities might be innate: languages would exploit the natural boundaries that coincide with phonemic borders and remove those that are linguistically irrelevant, through learning and exposure (Werker & Tees, 1984; Kuhl, 1993).

For other auditory stimuli, learning alone might be sufficient to create CP effects. Learning seems to be the explanation for why musicians show stronger CP for semitone boundaries (Siegel & Siegel, 1977) and musical intervals (Zatorre and Halpern, 1979) than non-musicians. Category learning has also been demonstrated in many laboratory experiments (Goldstone, 1994; Livingston, Andrews & Harnad, 1998; Guenther et al., 1999; see below).

1.1.3 Experimental procedure and definition

The typical experiments carried out to demonstrate categorical perception compare identification and discrimination performance. The identification test establishes the boundaries between categories. The stimuli are presented alone in a random order and the listeners are asked to label them. In the discrimination test the listeners are presented with pairs of stimuli and asked to tell whether they produced the same perceptual effect. There are different discrimination paradigms; all measuring the distance or the similarity among the stimuli (see Goldston, 1994b for a comparison between similarity and distance measures and Gerrits & Schouten, 2004 and for a critical review on the discrimination tasks).

According to Liberman's definition of categorical perception (Liberman et al., 1967), discrimination would be possible only for stimuli that are labeled differently and would drop to chance level for pairs of stimuli falling within one category. Experimental results are typically not in accordance with this definition, the size of the CP effect being purely a scaling factor: whenever a listener can better discriminate between sounds that lie on the opposite sides of a boundary than between sounds that fall within the same category (significant within-category compression and/or between-category separation), this is operationally defined as CP.

1.2. Critiques of the origin and interpretation of CP

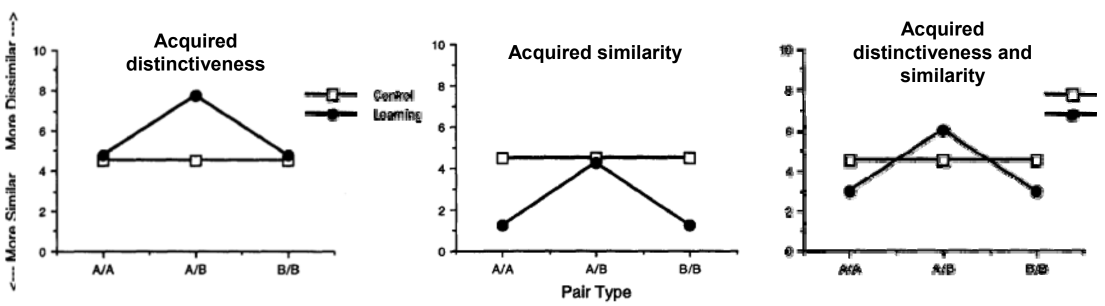
Some authors have challenged that CP is really a perceptual phenomenon, proposing that it might reflect a bias due to memory and task demands (Huttenlocher, Hedges & Vevea, 2000; Schouten, Gerrits, van Hessen, 2003; Gerrits & Schouten, 2004) or language influences (Roberson, Davies, Davidoff, 2000).

A memory account of CP assumes that whenever stimulus encoding is uncertain, there might be a shift in representation towards the category prototype (Huttenlocher, Hedges & Vevea, 2000). This is expected, for example when a task requires that the stimulus be held in memory for some time. According to this view, experimental conditions will affect the use of category information and this bias effect might induce CP. A similar position was proposed by Schouten (Schouten, Gerrits, van Hessen, 2003; Gerrits & Schouten, 2004; see also Pisoni, 1973), who pointed out that the kinds of discrimination tasks used in speech CP experiments promote the use of a labeling strategy in which sounds are first categorized into phonetic categories and the labels of the categories are compared.

Whatever its origin and interpretation, CP has been replicated in a variety of studies and supported by electrophysiological measures. Some of these studies are reviewed below.

1.3 CP development

The way in which CP effects might be modified or created by experience was first analyzed by Liberman (1961) for the case of phonemic boundaries. Liberman assumed that auditory perceptual space may undergo some sort of transformation, or “warping”, as a result of category learning, with members of the same category moving closer than they were before learning (Acquired Similarity) and/or with members belonging to different categories moving further than they were before learning (Acquired Distinctiveness). This might apply to entire dimensions, or to local regions within one dimension, that would get compressed or expanded as a result of learning. These learning hypotheses were adapted from developmental psychology (e.g. Lawrence, 1949; Nosofsky, 1986; Gibson, 1963; Goldstone, 1998).



Modified from Livingstone, Andrews & Harnad 1998

On the left: The first learning hypothesis, acquired distinctiveness (AD), states that discrimination of stimulus pairs is a priori generally poor. After category learning, the sensitivity for the stimuli around the category boundary increases and thus a discrimination peak forms.
In the middle: The second learning hypothesis, acquired similarity (AS), states that discrimination of stimulus pairs is generally good prior to learning and that it gets worse for stimuli that are categorized in the same class as a result of learning, so that only the category boundary remains discriminable.
On the right: Learning may be a combination of both: before learning, stimulus pairs are discriminated at an intermediate level; with learning, the phoneme boundary is sharpened through acquired distinctiveness, while discriminability within categories is reduced through acquired similarity.

FIGURE 1.2. Schema of the learning hypothesis proposed by Liberman. Modified from Livingstone, Andrews, Harnad, 1998.

1.3.1 Experimental procedure

The direct way to test how CP develops is to study the learning process by means of a baseline /training/ test paradigm.

A set of stimuli is tested, usually in pairs, for similarity or discriminability. In the case of similarity, a statistical technique called multidimensional scaling (Cox and Cox, 2001) is used to map the perceptual similarity of the tested stimuli. In the case of discriminability, same/different judgments and signal detection analysis are used to estimate the discriminability of a set of pairs. Then the same subjects or a different set are trained, using trial and error and corrective feedback, to group the stimuli into two or more categories. After the categorization has been learned, as assessed by an identification task, similarity or discriminability are tested again, and compared against the untrained data. Thus, the baseline for saying whether there was a warping of the space is given by the discrimination (or similarity judgment) before learning.

Category formation, with reference to Liberman's hypothesis of perceptual modifications, has been investigated in a variety of language studies. A brief review of these is given below. Only a minority of these studies addressed the learning process directly by comparing perceptual abilities before and after category acquisition. Recently a few training studies addressed perceptual learning related to the categorization. They are reviewed in more detail in Section 1.3.4.

1.3.2 Language studies

Four main lines of research have been employed in investigations of CP development:

- Comparison between language stimuli and non-speech controls.

In these studies (e.g. Liberman, 1961), the perception of speech sounds and non-speech controls that vary along the same continuum are compared. The discrimination of controls is taken as an indication of how the physical continuum is perceived before the phoneme category is formed, and the comparison between the discrimination function for the speech and non-speech sounds is considered to reflect the path of learning. These studies support the Acquired Distinctiveness (AD) hypothesis illustrated in Figure 1.2. However the correctness of the non-speech control conditions has been criticized (see Pisoni, 1991).

- Cross-language perception

These studies focus on non-native speech contrasts, as native phoneme categories are thought to be relatively fixed. CP development is addressed by comparing the identification and (sometimes) discrimination abilities for the nonnative contrast between groups of subjects that either are naïve to the contrast, or have had some experience with it while learning a second language, or have received laboratory training. The results for the nonnative speakers are compared to those of the native speakers as well. These data provide inconclusive evidence for both the Acquired Similarity (AS) and the Acquired Distinctiveness (AD) hypotheses.

For example, MacKain et al. (1981) compared the perception of the /r-l/ contrast between inexperienced and experienced Japanese users of English, and Americans. He found, in line with the AD hypothesis, that the inexperienced Japanese had difficulty discriminating any stimulus pair, while the experienced learners had a discrimination peak along the continuum near their perceived phoneme boundary, although less pronounced than native listeners. Similar results were found in several other studies (e.g. Lisker & Abramson, 1970; Miyawaki et al., 1975; Zahid, 1996).

On the other hand a similar study by Iverson et al. (2003) contrasts with MacKain et al. (1981) and might support both AD and AS. The perception of the /r-l/ contrast was compared for groups of speakers with different language background. Inexperienced Japanese users of English showed a homogeneous discrimination performance, lower than the Americans' discrimination level at its peak, but higher that the Americans' level for the within-category sounds. For the Japanese subjects to reach the native discrimination function a combination of AS and AD would be required, but the experiment does not show that they in fact learn to categorize /r/ and /l/ at the same level of discrimination performance that Americans do.

A possible confound in these experiments is that the level of experience with phonemic non-native contrasts is difficult or impossible to specify or to control for.

- Language learning in a laboratory

Some studies tried to change phoneme perception through laboratory training, but only a few of them measured both identification and discrimination performance before and after training to monitor any change in perceptual sensitivity that is related to classification (e.g. Strange & Dittman, 1984; Jamieson & Morosan, 1986; Tremblay, Kraus, Carrell & McGee, 1997; Heeren & Schouten, 2008). Not all training studies found CP for the phonemic contrast after training, although the improvement in classification was always salient (e.g. Heeren & Schouten, 2008). Other studies found that trained subjects develop a discrimination peak between the learned categories which, as in the case of second language learners, remains lower than the discrimination peak in native listeners. The results of these studies are in general consistent with the AD hypothesis.

- Language acquisition

CP formation has also been studied in the frame of language acquisition in infants and children. Infants up to six months of age are capable of distinguishing many of the world's speech sound contrasts, even if those contrasts do not occur in their own language environment (for a review, see Werker & Polka, 1993). By the age of 10-12 months, those same infants lose this ability (e.g. Werker and Tees, 1984, Kuhl et al., 2006, Kuhl et al., 2008). These results have been also confirmed by electrophysiology studies employing mismatch negativity (Cheour et al., 1998). The decline in sensitivity to nonnative phoneme contrasts has led to the conclusion that Acquired Similarity underlies CP development in speech (Pisoni, 1991), although it was later found that together with a decrease in the ability to discriminate non-native contrasts, native language phonetic differences increase (Kuhl et al., 2008), which would indicate that Acquired Distinctiveness is also involved.

However, to directly test any hypothesis, and compare studies on infants and adults, the research design needs to include a sensitive discrimination task producing high-resolution results. The youngest children that are thought to be capable of performing such a task are 6-year-olds (Hazan & Barrett, 2000). Studies on language learning in children from age five to fifteen show that the overall perceptual sensitivity to phonetic dimensions is generally lower in children than in adults and it increases with age (Sussman, 1993). Second language acquisition studies (e.g. Williams, 1979; Streeter & Landauer, 1976) found evidence that discrimination peaks near the category boundaries also develop with age and experience, which would be in line with the Acquired Distinctiveness hypothesis.

Summary of the language studies

Language studies seem to support the idea that CP for speech is acquired through exposure to and active use of language. Studies in children and adults are mostly consistent with the AD learning hypothesis, as discrimination peaks seem to form and refine with age and experience. Studies in infants are also consistent with the AS learning hypothesis, as infants in their first year of life seem to reduce their sensitivity for speech contrasts that are not relevant for their native language, while at the same time enhancing the relevant contrasts. Yet, it is difficult to compare language studies in infants with CP studies in adults, as the methodologies and experimental conditions are very different.

More generally, language experiments may not be able to provide a definitive answer to how CP develops, especially one that could be generalized to other stimuli and perceptual modalities, due to the peculiarity of speech categories. Even without regarding language as a special domain within the auditory modality, language categories are certainly expected to be over-learned, and precise measures of the effects of exposure and practice are difficult to obtain.

1.3.3 Non-speech learning studies

Direct evidence for CP development has been obtained in within-subject designs with non-speech stimuli. Such designs usually consist of a baseline measurement, training with the new materials, and post-training measurements, which are compared to the baseline. Within this paradigm, there is evidence that CP effects can be induced by learning alone (with visual stimuli: Lane, 1965, replicated and extended by Goldstone, 1994; Livingston, Andrews & Harnad, 1998; with non-speech acoustic stimuli: Guenther et al., 1999).

These studies found evidence supporting both types of learning, acquired similarity and acquired distinctiveness, probably induced by different stimuli and training paradigms. When more than one dimension is involved, the category-relevant dimension always got expanded, while the non-relevant dimension might be compressed, expanded or might show no change (Goldstone, 1994). Within the learned categories, however, results are contrasting, and both AD and AS have been found (Livingston et al., 1998; Guenther et al., 1999). These studies are briefly reviewed here.

Goldstone (1994) found evidence in support of both hypotheses, acquired distinctiveness and acquired similarity, but their occurrence depended on the identity and relevance of the stimulus dimension in learning. He used visual stimuli varying in both size and brightness or brightness and saturation level. In each condition, different groups of subjects were trained to categorize the stimuli according to each or both dimensions. A fourth group of subjects did not receive any training and was used as a control. Most of the analyses were actually run against this control rather than using a within-subjects design. Goldstone found that the stimulus dimension relevant for the categorization always got expanded compared to the control, while the irrelevant dimension could be either compressed (e.g. the size dimension for the brightness categorizers), or it could show no change at all (e.g. the brightness dimension for the size categorizers), or could even be expanded, though less than the relevant dimension (e.g. the brightness dimension for the saturation categorizers and the saturation dimension for the brightness categorizers). Goldstone explains this pattern of results as being specific to the stimuli and dimension: when two dimensions are orthogonal the expansion along the relevant dimension does not affect the irrelevant one, that could be either compressed or not; when the two dimensions are integral (and thus likely to interfere with each other and give rise to a global perception) it is more likely that the expansion effect on one dimension also reflects onto the second one. Regarding the effects of learning within the relevant dimension, the theory predicts AD at category boundary regions and AS within categories. Goldstone found that higher distinctiveness was always acquired along the entire relevant dimension, so that there never was any AS, but for some dimensions (size and brightness) the cross-boundary regions were locally more sensitized (the saturation dimension did not show a relevant difference).

Contrasting results were found by Livingston et al. (1998). Stimuli were complex, artificial or natural, visual stimuli differing along two dimensions. Within the relevant dimension, stimuli belonging to the same category acquired similarity, while no acquired distinctiveness was found for the region between the two categories (between-subjects design)   As Livingston et al. (1998) commented, the differences with Goldstone's study could be task-related, as Livingston used a similarity rating instead of a classification task, and/or stimulus-related.

A third study (Guenther et al., 1999) was performed with acoustic non-speech stimuli varying along a mono-dimensional continuum. Guenther used a within-subjects, baseline-training-posttest design and found that two different kinds of training on the same physical stimuli led to different perceptual effects. Categorization training, where subjects learned to identify sounds as belonging to a certain training region, led to a decrease of sensitivity to the stimuli within the trained category. Discrimination training (AX task) led to an acquired distinctiveness among the same stimuli.

It has to be noted, however, that the three groups of studies cited above used the same stimuli for training and testing, thus making it questionable whether an actual categorical distinction was learned, rather than memorizing the labels applied to each stimulus in the set.

In summary, previous evidence suggests that both Acquired Similarity and Acquired Distinctiveness may play a role in CP learning, with different studies documenting the involvement of either AD or AS. It has been proposed that when the objects being categorized are very different, then Acquired Similarities are more likely to be observed, while if they are very similar, Acquired Distinctiveness is more likely (Livingston et al., 1998).

1.4 Electrophysiology studies

Very little is known about the neuronal mechanisms supporting CP and governing category learning. Nevertheless, neuronal correlates of CP have been found in animal studies. For example primary auditory fibers of chinchillas show categorical encoding of VOT (Sinex and McDonald, 1989; Sinex et al., 1991; Sinex and Narayan, 1994) and sensorimotor neurons in nucleus HVC of swamp sparrows show categorical responses to changes in note duration corresponding to the birds behavior in perceptual experiments (Prather et al., 2009). In humans, imaging studies have shown a significant increase and decrease in cortical activity correlated with category learning (e.g. Reber, Stark, Squire, 1998).

Among the techniques adopted to probe CP in humans, mismatch negativity MMN (together with its magnetic counterpart MMNm) has proven to be particularly useful.

1.4.1 Mismatch Negativity (MMN)

MMN is a fronto-central component of the auditory ERP that reflects the detection of any discriminable change in a stream of identical auditory stimuli (for a review see Nätäänen & Winkler, 1999; and Nätäänen, 2001). MMN elicitation and amplitude have been shown to correlate with conscious sound discrimination performance (see Näätänen, 2001; for a case in which a discriminable change did not elicit MMN, see Shafer et al., 2004).

MMN is considered to be an 'automatic' measure of sensory pre-attentive processing because it can be elicited in the absence of attention. It is normally recorded during attentional distraction and it can be recorded during sleep (e.g. Sallinen et al., 1996), as well as in coma patients (e.g. Fischer et al., 1999, 2000) and in anesthetized animals (e.g. Kraus et al., 1994). MMN can be modulated by attention, and strongly focused attention away from the auditory stream may result in MMN suppression (Woldorff, et al., 1998).

MMN is measured in an oddball design protocol, where rare deviant sounds are embedded in a sequence of frequent standard sounds. It is a negative component appearing at around 100-250 ms after stimulus onset in the difference wave, obtained by subtracting the average response to the standard from the average response to the deviant.

The sources of the MMN have been located mainly in the auditory cortex (Alho, 1995) and in thalamic areas (Csepe, 1995). It has been suggested (e.g. Tiitinen et al., 1994; for a review see Näätänen, 1992 and Näätänen & Winkler, 1999) that MMN is the result of a memory comparison process detecting a discrepancy between the cortical neural representation of the repetitive sound features and the representation of the current stimulus. Stimulus-specific adaptation in auditory cortex has also been proposed as a possible mechanism underlying MMN (Ulanovsky & Nelken, 2003 and see also Jacobsen & Schroeger, 2001).

1.4.2 MMN and CP

In speech research, the MMN have been used to study categorical perception, but the results have been contradictory.

Some studies have found evidence that the MMN reflects phonetic categorical processing: the amplitude of MMN was found to be larger for pairs of stimuli crossing native phonetic boundaries than for within-category pairs (Aaltonen et al., 1997, Sharma & Dorman, 1999; Szymanski et al., 1999). Other studies report equally large MMN for the cross-category and within-category pairs of stimuli, suggesting that MMN might reflect acoustic similarities rather than the categorical perception of speech (Aaltonen et al., 87; Sams et al., 1990; Sharma et al., 1993; Maiste et al., 1995; Sharma & Dorman, 1998).

In studies that compare the perception of native and non-native phonemic contrasts, MMN has been found to be generally larger for the native contrasts (Näätänen et al., 1997), and to increase as a function of learning (Winkler et al., 1999). These results are in agreement with the behavioral results reviewed above. However, the MMN given in response to contrasts which are difficult to discriminate seems to be affected by attention (Garrido, Hisagi, Shafer, 2004; Hisagi, 2007), which may complicate the interpretation of MMN data in these contexts. MMN is also modulated by perceptual learning. This has been demonstrated by Näätänen et al., (1993) for tonal stimuli and by Kraus et al., (1995) for speech contrasts. In both studies, training to better discriminate stimuli resulted in larger MMNs to differences in those same stimuli.

In a laboratory categorization training experiment (Tremblay et al., 1997), identification training also resulted in a larger MMN for the cross-boundary pairs of stimuli after training, however no MMN was recorded for within-category stimuli, leaving the possibility open that training and exposure enhanced sensitivity to the stimuli in a non-category-specific manner.

1.5 Outline of this dissertation

The present thesis addresses two questions: first, whether new and reliable perceptual categories can be established for synthetic sounds varying along two dimensions (only one of which is relevant to categorization) through short-term categorization training; and second, whether such learning results in categorical perception, as assessed by two measures of perceptual sensitivity, discriminability and the MMN.

For this study a synthetic set of sounds was constructed and characterized with respect to naïve listeners' perceptions. Two non-overlapping structured categories were arbitrarily defined. A group of subjects, selected on the basis of their listening abilities, was trained to categorize the sounds. During training, categories were not defined explicitly: the stimuli were chosen randomly from the stimulus set according to the probability structure of the categories and the borders of the category had to be discovered through trial and error. Feedback was provided to guide learning, and the identification and discrimination abilities of subjects were measured before and after training.

The comparison between identification performance before and after training ensured that the categorization was learned appropriately. According to CP the way sounds are classified would determine the way they are discriminated, because of the distortions in the perceptual space that contains them. A comparison between discrimination performance before and after training, as assessed by a set of auditory staircases, was used to test this hypothesis.

To the same end, the MMN to pairs of stimuli straddling the category borders, or contained within the same category, was measured before and after training. As reviewed in Section 1.4, MMN is considered to be an index of discriminability, its amplitude scaling with the perceptual distance between the standard and the deviant. Thus, if category training induced CP, then the amplitude of the MMN would be expected to increase for between-category stimuli, signaling AD, and/or decrease for within-category pairs of stimuli, signaling AS. Thus, the goal of the electrophysiological part of this study is both to obtain objective measures of exposure- and learning-induced brain responses to the stimuli, as well as using this information to indicate whether category learning involves AD, AS or both.

Chapter 2

The present chapter documents the procedures that were used to create and select the final versions of the stimuli for the category learning experiment.

Language stimuli could not be used because of experimental design problems resulting from the over-learning and extensive daily experience that people have with phonetic categories. These problems also apply to non-native language contrasts, as they may often fall within the same acoustic space relevant for native language contrasts (there are rare exceptions such as click languages, which contain phonemes that are not perceived as language by non-native speakers). Linguistic stimuli are also acoustically complex, co-varying in many dimensions, so that it is not always straightforward to uniquely relate variation along a single physical dimension to variation along a single perceptual dimension.

An attempt was made to design a class of stimuli that listeners are naïve to, in the sense that subjects should have no specific categories already formed within the experimental perceptual space prior to laboratory training. Such sounds would not specifically remind subjects of familiar sound-producing objects. The idea was to have sounds that were “rich” enough in acoustic variation so that they would be acoustically engaging to the subjects, but at the same time could be easily perceptually manipulated by controlling variation in a few simple parameters. These considerations led to the design of the synthesized noise stimuli detailed below.

2.1 Generation of the sound stimuli

2.1.1 The perceptual attributes of sound

The basic subjective attributes of sound are loudness, pitch, and timbre.

In 1973 The American National Standards Institute gave a formal definition to these terms. Loudness is "...that intensive attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud", pitch is "…that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low", and timbre is "…that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different".

Pitch is the perceptual correlate of the periodicity in a sound. For pure tones this corresponds to the frequency, thus their pitch is unambiguous, although loudness may also influence pure-tone pitch percepts (Terhardt, 1974). Pure tones occur extremely rarely in nature. For complex natural sounds, where several frequencies are involved, pitch salience depends on the degree to which partials are harmonic. The pitch of harmonic sounds depends mostly on the frequency of the fundamental (for a review, see Handel, 1995; and Houtsma, 1997). The pitch of inharmonic complex sounds is usually less salient and depends on the frequencies of certain partials (Moore, Glasberg & Peters, 1985).

The timbre of a sound is itself a multidimensional attribute, depending on several physical variables. Many studies (e.g. Grey, 1977; McAdams & Cunibile, 1992) have investigated the relations between timbre perception and its underlying physical attributes, employing dissimilarity rating tests and analyzing these results with statistical mapping techniques such as multidimensional scaling (Cox and Cox, 2001). Quantitative treatments of timbre depend to some extent on the set of sounds under study, but in general three to four dimensions are enough to adequately describe the perceptual space of timbre. Timbre is influenced by a sound's frequency content, spectral profile, and temporal envelope (especially the initial "attack", including the synchronicity of onset of the sound's harmonics, but also by the decay and modulation of the steady-state portion). Although pitch and timbre have been defined as separate attributes, they depend on some of the same physical variables, and there can be perceptual interference between them (Melara & Marks, 1990a, 1990b, 1990c; Krumhansl & Iverson, 1992).

Pitch and timbre play an important role in the perception of speech and music. In speech, the pitch contour of a spoken sentence is the principal carrier of prosodic information (in tone languages, the pitch contour also conveys semantic information). Variation in the timbre of speech sounds, although not commonly described this way in the speech literature, enables listeners to identify phonemes and vowels necessary for speech segmentation. Pitch is determined by the rate of opening and closing of the vocal folds (glottal-pulse-rate, GPR), and timbre depends physically on the shape of the glottal air flow pulse and the instantaneous shape and length of the vocal tract. Both pitch and timbre are used in recognition of speaker gender and body size (Smith & Patterson, 2005). In music, pitch allows listeners to follow the contour of a melody, while the sounds of different auditory objects such as particular musical instruments can be separated on the basis of perceived timbre (see Handel, 1995).

2.1.2 Stimuli

The sound stimuli made for this study vary along two dimensions (pitch and timbre), but only one of these dimensions (pitch) was relevant for categorization. The second dimension provides additional ecological validity to the stimuli (natural sounds vary in irrelevant dimensions) as well as extra variability (a variable set of stimuli has been shown to facilitate learning of categories in a way that generalizes across the training stimuli according to Lively et al., 1993).

The stimuli were synthetically constructed using the SIGNAL Digital Programming Language (version 5.04.17, Engineering Design, Berkeley, CA). Two sets of stimuli were prepared for the study and one was finally chosen based on the lower overall confusion rates between pitch and timbre that it produced in the pilot discrimination experiments described below (Section 2.5).

Both sets of sounds consist of white noise filtered with a spectrum made up from the sum of 10 independent filters. The first of these filters was centered at the fundamental frequency, and the remaining nine were centered at integer multiples of the fundamental frequency from 2 to 9. Each filter had a Gaussian shape, and the width of each filter was small enough (50 Hz full width, half maximum) to produce a salient pitch related to the fundamental frequency, f_0. Timbres were varied either by changing the relative amplitudes of a set of filters of constant width, or by changing the width of the filters. If the width of the filters increases beyond around 130 Hz, the "tonal" pitch quality is lost and a diffuse noise is heard. Most of the stimuli used here had a "tonal" quality, but in one of the two stimulus sets the timbre was varied from "tonal" to "noiselike" in order to help subjects understand the nature of variation along this dimension.

The timbre dimensions were varied differently for the two sets:

1. The first set, called the "Gaussian" or "G set", was obtained by changing the relation between the amplitudes of the filters in the bank. The amplitudes always decayed from the first to the tenth filters, but the slope of decay varied from shallow to steep. Very generically, these sounds could be described as going from a sharp to a dull timbre as the higher frequency components experienced a greater decay.

G set

$$A_{ik}(f) = S_{ik} \cdot F_i(f) \qquad S_{ik} = \begin{cases} e^{-\frac{(1-i)^2}{2\cdot p_k^2}} & \text{if this quantity is bigger than 0.036} \\ 0.036 & \text{otherwise} \end{cases} \qquad F_i = e^{-\frac{(f - f_0 \cdot i)^2}{2\cdot\sigma^2}}$$

FORMULA 2.1. "G set". Each filter function $A_{ik}(f)$ is a Gaussian function $F_i(f)$ weighted by a scaling factor S_{ik}. The amplitude of the filters was scaled in a way that the lower filters would be no more than 15 dB_V smaller than the first filter. σ = 20 Hz; f_0 = fundamental frequency. i = 1,2,3,…10 indicates the filter number; k = 1,2,…9 indicates the sound within the G set. At the beginning of the study p_k = 0.5,1,1.5,… 4.5; then it was adjusted based on the perceptual timbre distances (see Section 2.4).

2. The second set, called the "Noisy set" or "N set", was obtained by choosing one exponential decay function (thus leaving the relative amplitudes of the filters constant) and changing the width of each Gaussian filter from narrower to wider. The timbre quality of these sounds could be described as going from "tonal" to "noiselike". The sounds in this set may be thought as varying in the saliency of their pitch.

N set

$$A_{ik}(f) = S_i \cdot F_{ik}(f) \qquad S_i = e^{\frac{(1-i)}{2}} \qquad F_{ik} = e^{-\frac{(f - f_0 \cdot i)^2}{2\cdot\sigma_k^2}}$$

FORMULA 2.2. "N set". Each filter function $A_{ik}(f)$ is a Gaussian function $F_{ik}(f)$ weighted by a scaling factor S_i. f_0 = fundamental frequency; i = 1,2,3,…10 indicates the filter number; k = 1,2,…10 indicates the sound within the N set. At the beginning of the study σ_k = 10,15,…55 Hz, then it was adjusted based on the perceptual timbre distances (see Section 2.4).

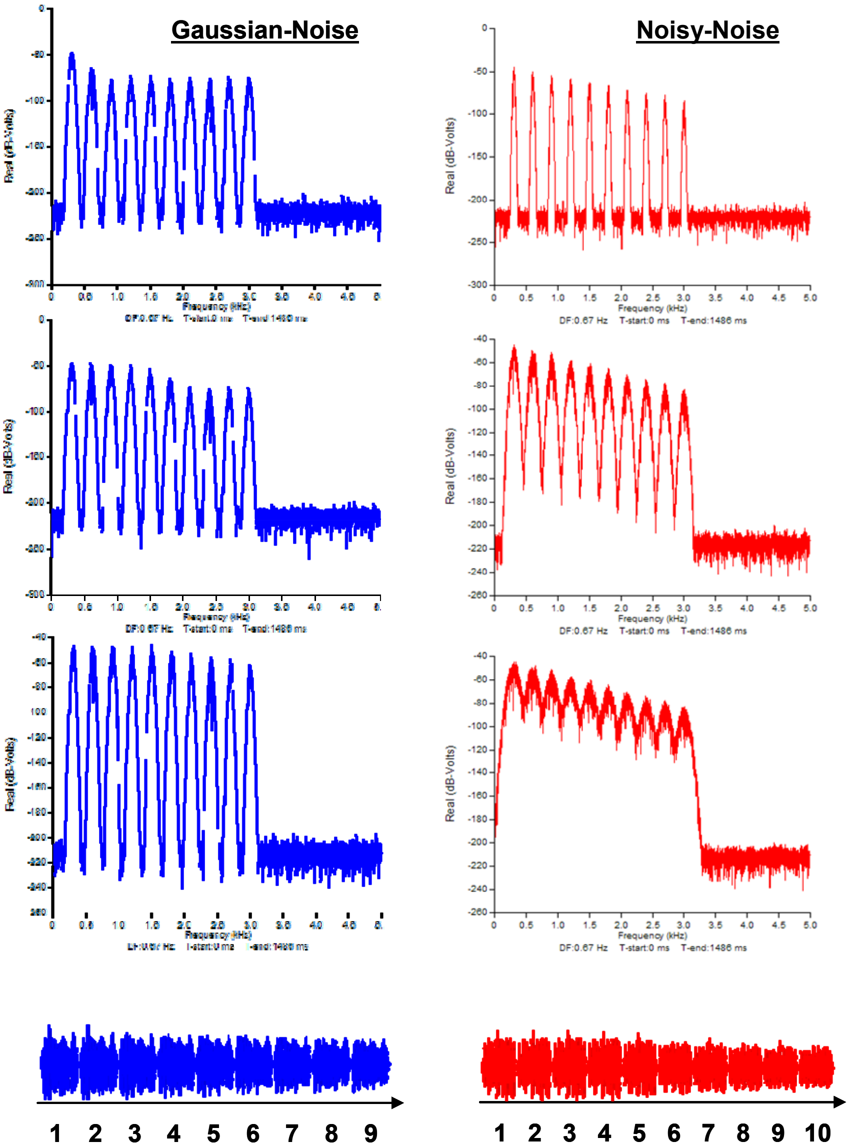


FIGURE 2.1: Above: on the left, in blue, spectral shape of three sounds from the "G set" (stimuli 1,5,9 from the series); on the right, in red, spectral shape of three sounds from the "N set" (stimuli 1,5,10 from the series). Below: amplitude-time representation of the stimulus series for the two sets of sounds.

2.2 Tests on the stimuli

Perceptual experiments were conducted in order to assess how naïve listeners perceived these sounds. Since pitch and timbre percepts may interfere with each other; it was important to find stimuli which minimized this interference, so that categories based on variation in one of these attributes could be clearly perceived by subjects in spite of variation in the other attribute. Two tests were run with both sets of stimuli:

1. a rating task to scale the similarity of the sounds along the timbre dimension
2. a test on pitch-timbre interference, to assess which of the sets allowed the best separation between timbre and pitch

2.3 Rating task

A dissimilarity rating task was used to test how naïve listeners perceived the timbre of the sounds. The goal of this procedure was to adjust the parameters in the sound generating function in order to obtain a set of stimuli evenly spaced along the timbre dimension. This procedure was not required for pitch, because it is already known that its perception is log-linear – pitch distances depend on frequency ratios between the stimuli (Wier et al., 1977).

Nine sounds were taken from the Gaussian set, with nine different levels of timbre and one level of pitch, and these sounds were presented in all possible pairs (9x9). In the same way, ten sounds were taken from the Noisy set, with ten different levels of timbre and one level of pitch, and these were also presented in all possible pairs (10x10). The presentation conditions are described in more detail below. A group of subjects were tested with both sets of sounds in two different sessions separated by at least one day. They were asked to rate how different the sounds in each paired presentation were, on a scale from 0 to 9, where 0 means "identical" and 9 means "very different". They were not instructed about, and not asked to focus on, the way in which the sounds vary, but were only asked to judge how similar the sounds were. For each subject, the mean over six ratings of the same pair was computed. The data from two subjects were excluded, each in one condition only, because their ratings did not appear to scale with parametric distance (see below). The dissimilarity judgments were averaged across subjects and processed using multidimensional scaling (Cox and Cox, 2001) to reconstruct a map of the position of the sounds in an acoustic space that could best account for the perceived distances. This representation was used to modify the parameters of the sound generating functions in order to obtain even perceptual distances between any two adjacent sounds in a set.

Multidimensional scaling (MDS)

MDS models assume the existence of an underlying multidimensional space that describes a set of items represented as points in that space. The objective of multidimensional scaling is to characterize (extract) the dimensions of the space and to describe the position of the items in it, starting from a matrix of perceived distances among the items. The matrix of distances is usually made up from people's judgments concerning the degree of similarity of pairs of stimuli, measured either as metric variables (interval or ratio scaled), or as non-metric variables (ordinally scaled).

In this experiment the items are the sounds from the two sets, the space is the perceptual auditory space and the judgments used to build the distance matrix are the similarities between all possible couples of sounds, as assessed by the subjects in the rating experiment. Since the response scale used here only assumes an ordinal scale of measurement, a non-metric MDS algorithm was used. The results of the analysis provide a quantitative representation (map) of the sounds in the acoustic space. The dimensions of the acoustic space represent attributes that are perceived as characterizing the sounds. As all sounds were equated for duration, intensity and pitch, it is assumed that the dimensions found by the MDS algorithm are inherent to (ascribed to) timbre. Note that the goal of this analysis is not to try to interpret what acoustic variables contribute to timbre, but rather to have a systematic way of reorganizing the parameters of the sound-generating functions in order to produce a set of stimuli that a typical human subject will perceive as evenly-spaced from the point of view of timbre.

2.3.1 Materials and methods

Subjects: 8 subjects (2 males and 6 females, age 21-35 years) were tested with both sets of sounds in two different sessions separated by at least one day and run in a random order. These participants were randomly chosen from volunteers recruited by telephone. They all gave informed consent and reported having no hearing deficits. They were paid 8 € per hour.

Stimuli: Sound stimuli were based on modifications of white noise as described in Section 2.1.2. A spectrum consisting of a sum of ten Gaussian filters centered at the first 10 multiples of a fundamental frequency f_0 = 300 Hz was convoluted with white noise to produce each stimulus. Within each stimulus, the ten filters have the same width, while their amplitude decreased as described below. Pitch in these sounds is related to their fundamental frequency f_0, while timbre is related to the spectral profile of the sounds. “G set” sounds had the width of the filters kept constant (full width at half maximum of each filter was around 50 Hz) for all stimuli, while the relative amplitude of the 10 filter peaks changed from one stimulus to the next according to Formula 2.1 (see also Figure 2.1). There were nine sounds in this set, yielding 81 pairs of similarity judgments between all sounds. “Nset” sounds had the amplitude ratios of the 10 filter peaks held constant, while the width of all ten filters changed linearly by the same relative amount for each sound, according to Formula 2.2 (see also Figure 2.1). There were ten sounds in this set, yielding 100 pairs of similarity judgments.

Each sound lasted 500 ms including 50 ms of a cosine-shaped rise and fall at the sound edges. To obtain the same perceived loudness the sounds were equated for their RMS value (Soloudre & Norcross 2003; Soloudre 2004).

Subjects made judgments on each complete set of stimulus pairs in randomized order; these judgments were repeated 3 times with brief resting breaks (~ 5 minutes, at the discretion of the subject) in between each complete set.

Procedure: The test took place in a quiet room. Subjects were seated in front of a computer monitor and keyboard. The sounds were amplified and delivered binaurally through Sennheiser HD 580 headphones. Responses were indicated by key presses on the keyboard. Subjects were instructed to choose a comfortable level of sound (which they could adjust by rotating a knob) at the beginning of the experiment and to leave the level the same throughout the entire test series.

The session started with an example of each of the sounds in the set under test (listening to the examples lasted around 15 seconds). Then, subjects were presented with pairs of sounds and asked to rate how dissimilar the two sounds in the pair were on a scale from 0 to 9, where 0 means “identical” and 9 means “very different”. Each pair could be heard more than once.

The time course of each trial was as follows. The two sounds in a pair (500 ms each) were played with an interval of 300 ms of silence between them. Then, after 500 ms, a popup window appeared on the monitor asking the subjects to rate the pair or listen to it again. There was no time limit for providing an answer. After the response was made there was an interval of 500 ms before the next trial started (Figure 2.2)

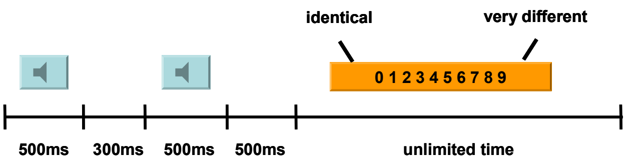


FIGURE 2.2: Time course of one trial in the rating task.

2.3.2 Analyses

All analyses were performed using MATLAB R2007a (Version 7.4.0.287, The MathWorks, Natick MA). For each subject (N=8), the mean dissimilarity rating for stimulus pairs with a parametric distance equal to one step in the series was computed. There were 8x6 = 48 such pairs for the G set and 9x6 = 54 for the N set. The slope of the two regression lines were significantly different (F = 6.67, df =1; p = 0.0109), indicating that the two sets differed in the degree to which the average perceptual distance of adjacent stimuli changed throughout each stimulus series. The N set exhibited more consistent (flatter) ratings of perceptual similarity across the series than the G set did (Figure 2.3).

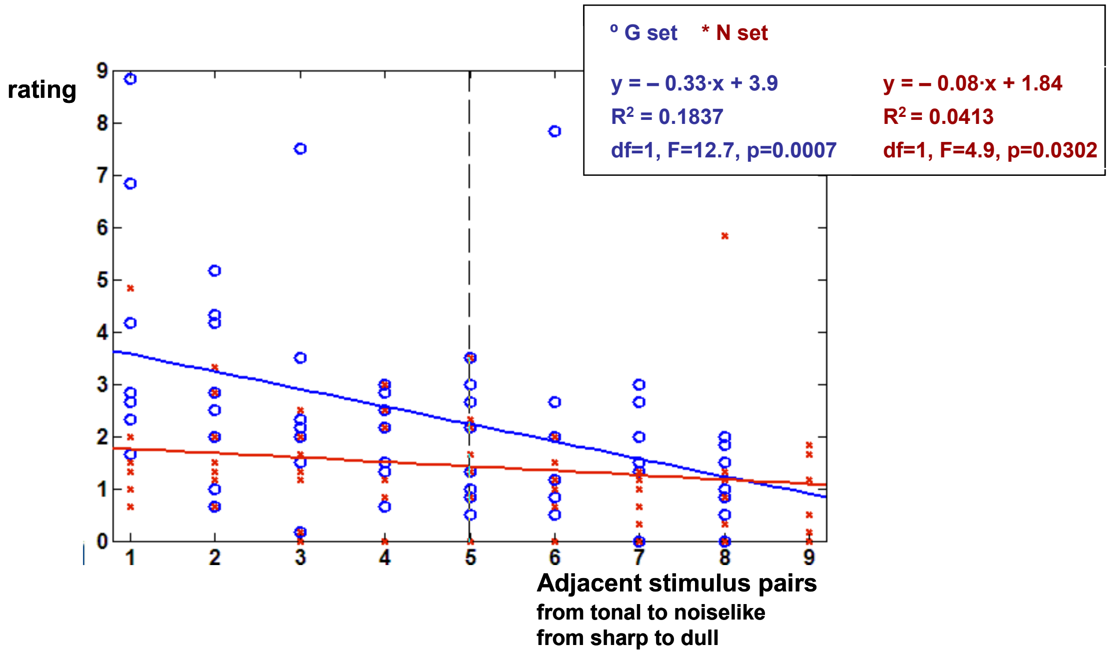


FIGURE 2.3. Dissimilarity ratings linearly regressed over stimulus pairs with parametric distance of one step (“adjacent”) for the two sets. Same 8 subjects in both conditions.

MDS analysis was carried out after arranging the dissimilarity judgments for each subject into a 10 x 10 (N set) or a 9 x 9 (G set) upper triangular matrix of data (no diagonal). Each non-zero cell in the matrices contained the average of six ratings of the same pair of sounds (pairs made up of the same two sounds presented in a different order counted as the same pair).

These ratings were first grouped according to the timbre parametric distances and a linear regression of the average ratings of each subject on the parametric distances was computed (Figure 2.4).

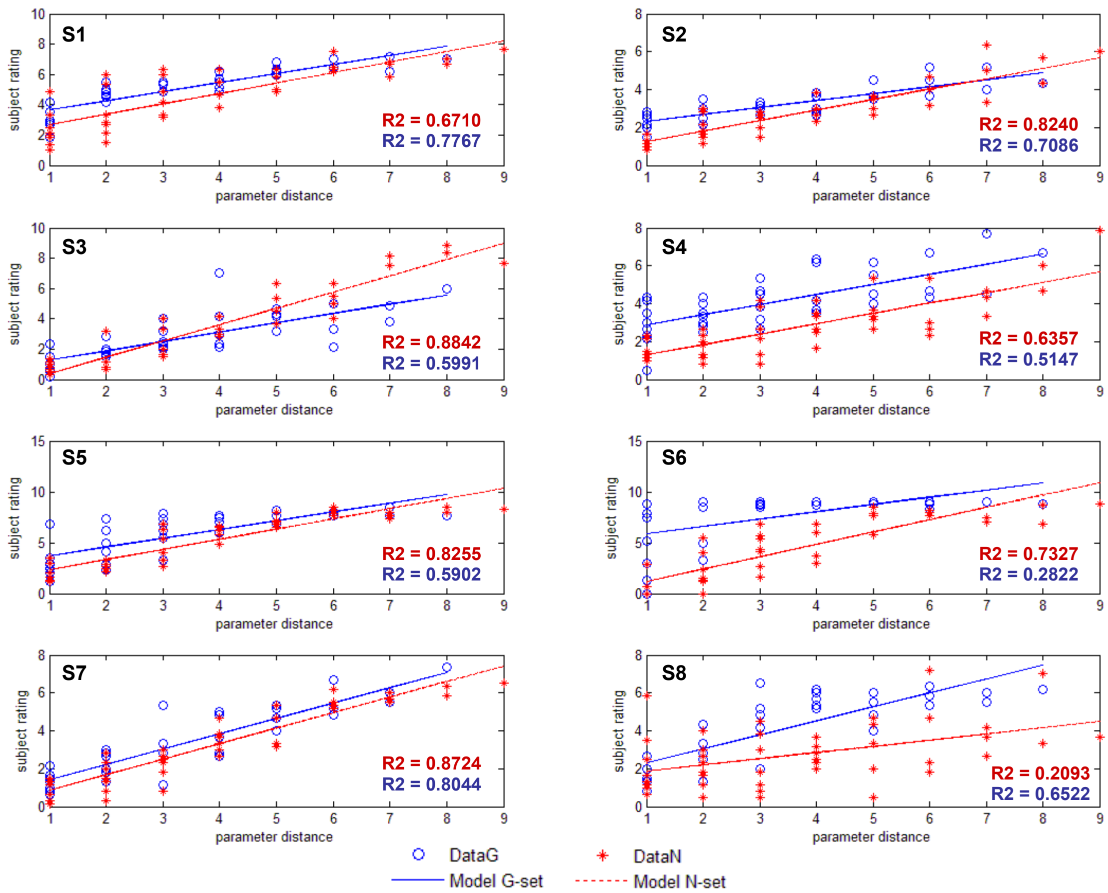


FIGURE 2.4. Average ratings of each subject regressed on the parametric distances. All p-values for the regressions were significant at $p < 0.05$.

Subjects with plots in which r^2 was smaller than 0.3 were considered to provide unreliable data and were excluded from the analysis. There were two exclusions: subjects 6 (female, r^2 = 0.2822) in the G set and subject 8 (male, r^2 = 0.2093) in the N set.

The triangular matrices of dissimilarities for all the remaining subjects (7 for each condition) were averaged across condition. The two resulting grand mean dissimilarity matrices were processed by non-metric multidimensional scaling analysis. Spatial representations obtained for three dimensions are shown in Figure 2.5 on the left. Figure 2.5 (right) shows the values of "stress" associated with the MDS results in different dimensions. Stress is the most common measure used to evaluate how well a particular configuration reproduces the observed distance matrix (Cox & Cox, 1994, 2001). In both conditions, the stress was small (less than 0.1), for the one-dimensional solution and it was almost zero at 4 or more dimensions.

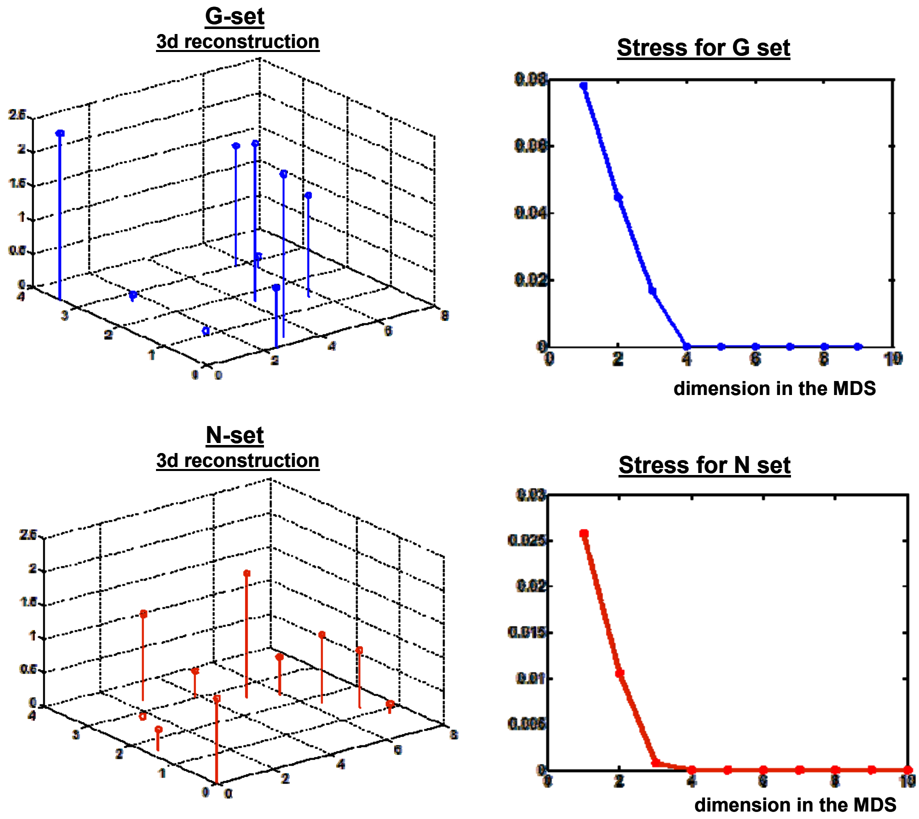


FIGURE 2.5. On the left, three-dimensional spatial representation based on the dissimilarity ratings for nine sounds from the G set and ten sounds from the N set, On the right, stress as a function of the dimensionality of the solution.

2.3.3 Results

In comparison with the G set, the N set was rated more homogeneously along the timbre dimension, as indicated by the regression analysis of the adjacent pairs (Figure 2.3). The MDS analysis suggested that the timbre of the tested sounds was best described by a two or three dimensional space for the N set and by a three or four dimensional space for the G set. The MDS solutions in a three dimensional space were used to adjust the parameters of the sound generating function, in order to obtain homogeneous step sizes, and a new timbre series was produced.

2.4 New timbre series

For constructing the new series, distances between adjacent parameters (Δk_{ij}) were proportional to the inverse of the perceived distances (Δp_{ij}) between adjacent sounds, as reconstructed by the MDS. They were computed as follows: first the inverse of all the perceived distances (Δp_{ij}) between adjacent sounds were computed. Then, these values were multiplied by a constant to make the new parameter range (Δk_{N1}) equal to the old parameter range (Δh_{N1}).

After this adjustment, two new sets of stimuli that were more equally-spaced with regard to timbre were produced and the interference between pitch and timbre within these sets was evaluated as described below.

2.5 Discrimination listening task

Pitch-timbre interference was studied using a discrimination listening task. A new group of subjects was asked to focus on the dimension of timbre while listening to couples of sounds that varied in both pitch and timbre. Their task was to judge whether the two sounds in a couple were identical along the timbre dimension.

The sounds used for this experiment were generated in the same way as those used in the rating task, but the parameters of the filters were adjusted according to the results of the rating task as described above, in order to produce two sets that varied more homogeneously along both the timbre and the pitch dimensions.

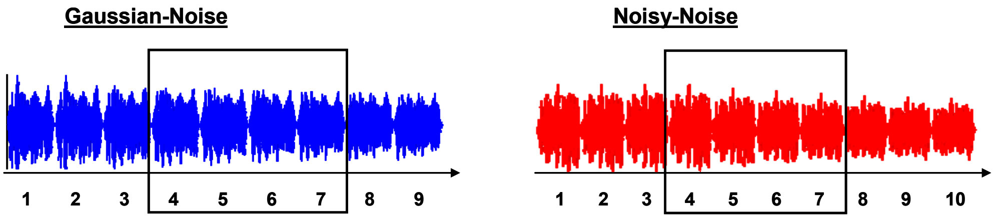
In everyday life people attend to both pitch and timbre and use them in order to extract information from sounds. An example is listening to music: we mostly use pitch to follow a tune and timbre to distinguish different instruments in an orchestra. Nevertheless, it is not necessarily very easy to selectively attend and judge one of the two dimensions on command, especially when one is inexperienced at formally describing sound percepts. For these reasons, this procedure contained an initial training phase where subjects could practice their selective listening skills for these sound attributes.

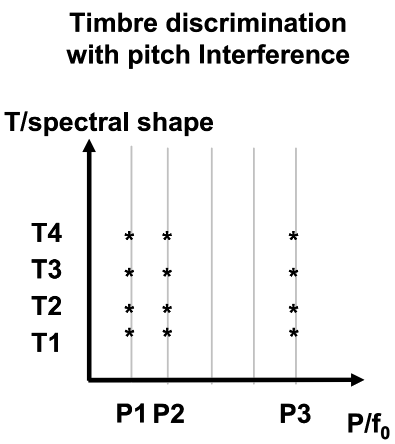
2.5.1 Materials and methods

Subjects: Nine subjects (1 male and 8 female, age 19-35 years) were tested with both sets of stimuli in two different sessions separated by at least one day. They were randomly chosen from volunteers recruited by telephone. They all gave informed consent and reported having no hearing deficits. They were paid 8 € per hour for the participation in the experiment. None of the subjects in this experiment had participated in the previous rating experiment.

Stimuli: The sound stimuli were produced as described in Section 2.1.2, with the exception of the modifications to the spectra described in Section 2.4.

Only four levels of timbre and three levels of pitch were used (see Figure 2.6). For the timbre dimension, four adjacent sounds one spectral step apart were used; for the pitch dimension two sounds spaced one semitone apart, and one sound three semitones away from its nearest partner (300, 318, 378 Hz) were chosen. The 12 sounds were combined in all possible pairs (144).





ΔP / ΔT	*0*	*1*	*3*	*4*
0	N=12	N=8	N=8	N=8
1	N=18	N=12	N=12	N=12
2	N=12	N=8	N=8	N=8
3	N=6	N=4	N=4	N=4

FIGURE 2.6. Upper plots: 4 central timbre levels were used in the interference task. Lower plot, left: schematic representation of the stimuli used in the test in the pitch-timbre space. Lower plot, right: table indicating, out of the 144 possible pairs, the number of pairs for each combination of the timbre and pitch distances. In the test phase this matrix is repeated three times.

Procedure: Testing procedures were identical to those described for the rating task (Section 2.3.1), except as explained below.

The experiment was divided into a training phase and a test phase.

1. **Training phase.** The training session started with a simple explanation of what pitch and timbre are. Subjects were asked to focus their attention on timbre and try to ignore as much as possible any pitch variation. Four examples of pairs were presented. After the examples the training began: subjects were presented with all the pairs of sounds in a random order and asked to answer the question “Do these two sounds have the same timbre?” The sounds could be heard more than once. After each response subjects received feedback in the form of a message in a popup window. The message stayed on screen for 1.3 seconds and could be one of two phrases: “Well, done! Go to next trial” or “Oh, no! That was wrong! Try again”. Embedded within the training procedure were little tests that the subjects undertook whenever they gave five correct answers in a row during training. In these tests the subjects had to judge ten pairs from among the most difficult pairs (those that have a close but not equal timbre). There was no feedback. If the outcome of the test was 7/10 or more correct

answers, the subjects moved on to the test phase. If the results were less than 7/10 correct responses, the subject continued the training phase. The training phase could only be stopped if a test were passed or if a limit of 288 practice trials were reached (in which case the subjects were not retained for the testing phase). One subject out of nine was rejected for the G series tests, three subjects out of nine were rejected for the N series tests. Subjects were allowed a break of around 5 minutes after their first 144 trials.

2. **Test phase.** All the possible pairs made up from the 12 initial sounds were presented for discrimination in random order. For each pair subjects were asked the question “Do these two sounds have the same timbre?”. There was no time limit for responding and no feedback. Overall there were 3x144 trials, allowing six repetitions for each pair of sounds (pairs made of the same two sounds presented in a different order count as the same pair). Two small breaks of around 5 minutes each were allowed.

The time course of each trial is shown in Figure 2.7. The two sounds in a pair (500 ms each) were played with an interval of 300 ms of silence between them. Then, after 500 ms a popup window appeared, asking the subject to make the discrimination. There was no time limit for answering. After a response was made, a feedback message appeared for 1.3 seconds in the training task only (in the test phase the intertrial interval started directly after the response). There was an interval of 500 ms before the next trial started.

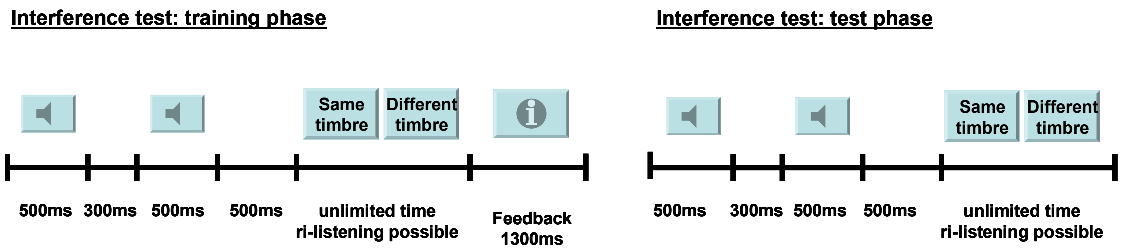


FIGURE 2.7. Time course of one trial in the interference test for the training phase (on the left) and the test phase (on the right).

2.5.2 Analysis

Of the nine subjects (8 female), eight (7 female) were able to complete the test for the G set, and six (5 female) were able to complete the test for the N set. Five subjects (4 female) were able to complete both sessions of the experiment.

In the training phase, a Wilcoxon rank-sum test (equivalent to a Mann−Whitney U test) revealed that there was no significant difference in the number of trials needed to complete training for the two sets of sounds (ranksum = 60, n1 = 8, n2 = 6, p = 0.0546), however the results were close to significance (with fewer trials for the G set). In the test phase, performance was significantly higher for N set sounds than for G set sounds (83 vs. 76% mean correct responses across subjects, Wilcoxon rank sum test (ranksum = 61, n1 = 8, n2 = 6, p = 0.0426).

Each pair of stimuli could have a parametric distance in pitch and timbre of 0, 1, 2, 3, steps for timbre and 0, 1, 3, 4 steps (semitones) for pitch. The pairs of sounds were grouped according to their pitch and timbre distance, following the schema of Figure 2.6. For each single subject, the proportion of “different timbre” answers was averaged for the six repetitions of the stimulus pairs in each group. When parametric distance in timbre was different from zero this proportion corresponded to accuracy; for zero parametric timbre distances it corresponded to one minus accuracy. Figures 2.8 and 2.9 show performance as a function of distance for each set of stimuli. Pitch interference on the percentage of “different timbre” responses was analyzed in the critical cases where timbre parametric distances are equal to zero or one. A Two-Way ANOVA with factors pitch distances (four levels) and timbre distances (two levels) was run separately for the two conditions (see Figures 2.8 and 2.9 on the left).

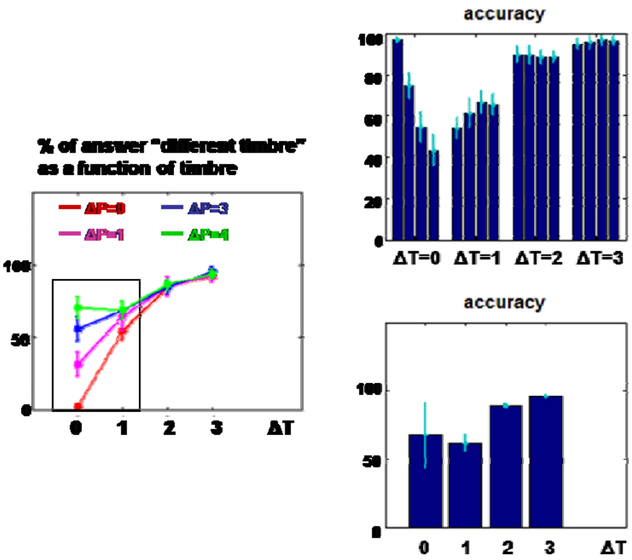


FIGURE 2.8. Results for the G set. Left plot: mean group proportion of “different timbre” answers and variability across subjects as a function of timbre distances, for different values of ΔP. On the right: mean group accuracy and variability as a function of timbre (upper right panel: ΔP increase in the direction left-to-right; lower right panel: accuracy averaged across different ΔP).

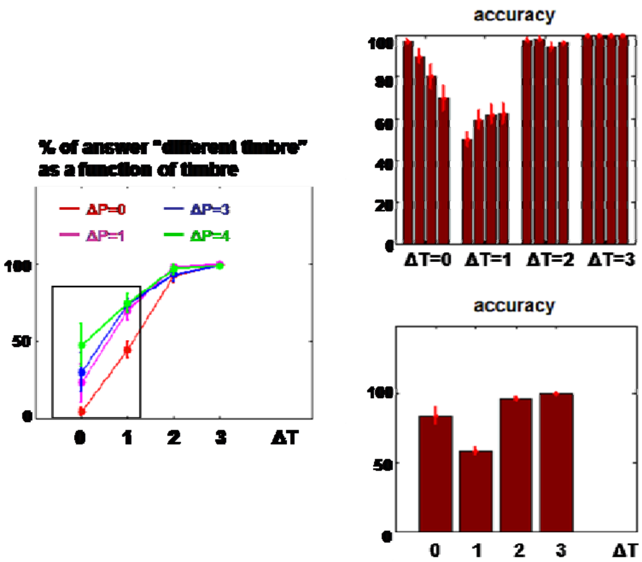


FIGURE 2.9: Results for the N set. Left plot: mean group proportion of “different timbre” answers and variability across subjects as a function of timbre distances, for different values of ΔP. On the right: mean group accuracy and variability as a function of timbre (upper right panel: ΔP increase in the direction left-to-right; lower right panel: accuracy averaged across different ΔP).

For the G set there was a significant interaction between pitch and timbre (N = 8, F=5.17, df=3, p=0.0032). For the N set there was no interaction, but a main effect of both timbre and pitch distances (N=6; for timbre F=38.73, df=1, p<0.0001; for pitch F=6.04, df =3, p=0.0017).

The average accuracy of subjects for pairs with no timbre variations was regressed against pitch variations for the two sets (Figure 2.10). The slopes were not significantly different (F = 3.03, df =1; p = 0.0878), suggesting that subjects showed a similar overall pattern of pitch-timbre interference for both sets of stimuli, even though the interference was less marked for the N set.

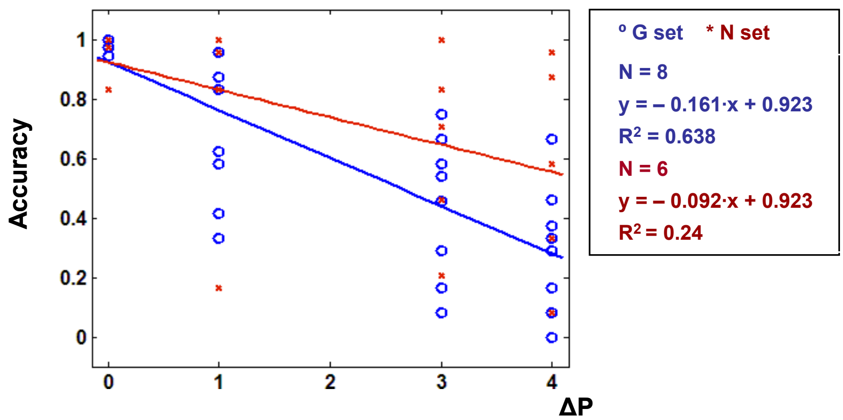


FIGURE 2.10. Subjects’ accuracy linearly regressed against pitch variations, when timbre is held constant.

2.5.3 Results and Discussion

The sounds from the two sets were discriminated in a parallel way. When timbre distance was 2 or 3 steps, accuracy was almost 100% for all subjects in both conditions, irrespective of pitch variations; for small timbre distances the performance decreased and was slightly influenced by pitch; for zero timbre differences the pitch influence is maximum in both sets, suggesting that subjects could not perfectly distinguish between small timbre and pitch variations. A comparison of the pitch influence in the two sets when timbre level was kept the same revealed that there was no clear advantage of one set over the other. Small pitch variations confounded the ability to judge timbre dissimilarities in both sets of sounds.

While the results from the discrimination test did not clearly point to one set of sounds being easier for normal subjects to attend to during selective listening, a comparison of Figures 2.8 and 2.9 and 2.10 suggests that the N set displays a smaller confounding effect than the G set. The discrimination test also showed that these sounds were perceived as being more similar to one another along the timbre scale. For these reasons, N-set stimuli with a timbre step size equal to twice the step of the series used for the discrimination task were selected for use in the categorization experiment.

Finally, the discrimination test confirmed the expectation that selective attention and discrimination of timbre variations in the presence of pitch variations might be very difficult for some subjects. Indeed, out of nine subjects, four failed to pass the training phase for one of the two sets, and the performance of those who succeeded was still affected by pitch variation. Since pitch and timbre play an important role in training subjects for the categorization experiment, the decision was made to set criteria for subject inclusion in the categorization experiment based on a minimum ability for selective perceptual separation between pitch and timbre. An entrance test for the mutual interference of pitch and timbre (a variation of the discrimination test reported above and described in the next chapter) was subsequently designed for this purpose.

2.6 Summary

This chapter discussed the design of the stimuli and the procedure for selecting the final stimulus set used for the category learning experiments. The goal was to obtain synthetic stimuli varying along the ecologically relevant dimensions of pitch and timbre (both central to the perception of music and speech). Starting from white noise filtered with a spectrum that was the sum of ten Gaussian filters, the filter parameters were varied to obtain two sets of sounds. Pitch in these sounds was varied by changing the central frequency of the filters. Timbre was varied by changing the spectral profile of the sounds (both in terms of filter shape and the relative amplitudes of the output of each filter). Two considerations needed to be addressed heuristically to select the final stimuli: the linearity of timbre perception and the degree of interference between pitch and timbre. Two psychophysical procedures were run to assess these issues, using a dissimilarity rating task on stimuli that varied only along the timbre dimension to assess timbre linearity, and a discrimination task on stimuli that varied along both dimensions to assess interference.

The rating task showed that both sets had nonlinear timbre variation. The results of this experiment were used to adjust the stimulus synthesis parameters in order to space the stimuli more evenly in timbre perceptual space.

The discrimination test, run after the adjustment of the timbre parameters, indicated that one type of stimulus sound (the N set) yielded a slightly better separation between timbre and pitch after training. Thus, these sounds were chosen as the stimuli for the subsequent categorization learning experiment.

Chapter 3

According to the categorical perception theory, classification determines discrimination. If one started with a perceptual space that was a homogeneous continuum (such as an auditory space that goes from low to high frequencies), then category learning would deform this space, so that sounds within the same category would be more similar to each other than sounds that are the same distance apart on the original continuum, but belong to different categories.

This scenario was reproduced and tested here by creating two artificial, non-overlapping, structured categories on the dimension of pitch, which is continuous on a logarithmic scale. A group of subjects selected on the basis of their listening abilities (see below) were trained to categorize sounds according to the two artificial categories.

Before and after training, the ability to assign stimuli to the categories was assessed by identification tests. The effects of categorization learning on discrimination abilities were assessed by two different means: a set of discrimination threshold tests (auditory staircase procedures) centered at various levels along the pitch dimension, and electroencephalographic (EEG) recordings of “oddball” stimulus detection for four sounds that occupied different positions along the pitch dimension.

This chapter describes the methods used for the categorization experiment. A schema of the experimental design is shown in Figure 3.1. The chapter is divided into two parts: the first part describes the selection of subjects and the second part describes the categorization procedure, task by task. The analyses and results are described in the next chapter.

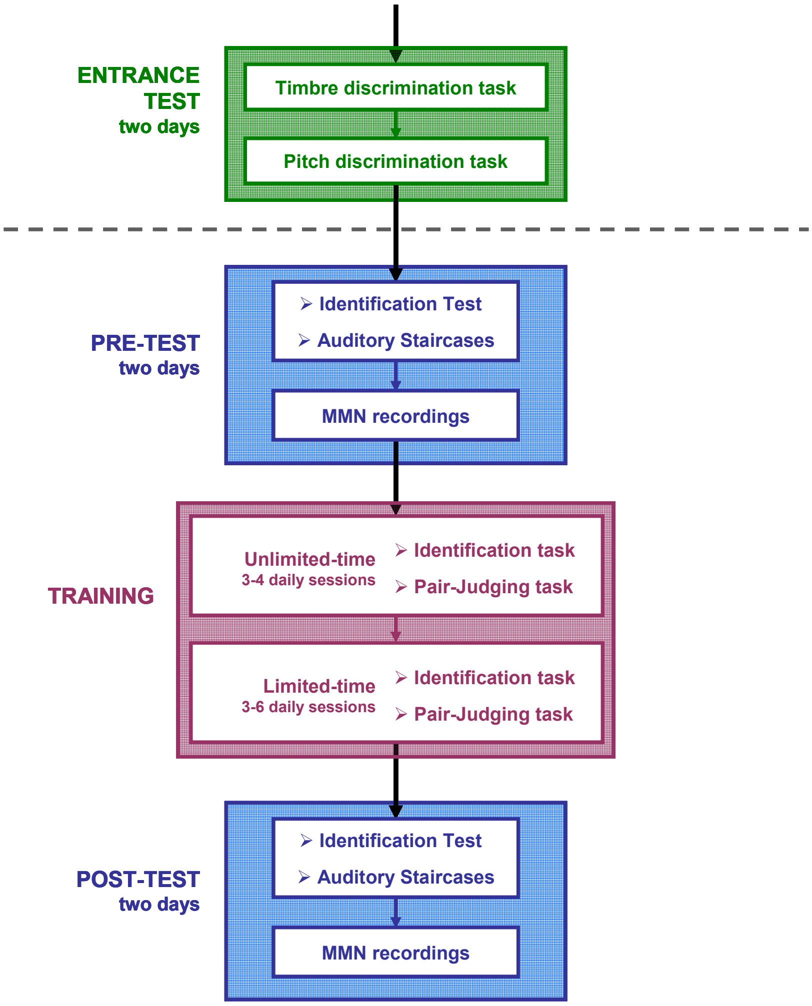


FIGURE 3.1. Tasks used in the categorization experiment.

3.1 Subject selection procedure: timbre and pitch discrimination tasks

As documented in Chapter 2, pitch-timbre interference was present in the experimental sounds, and in some subjects the interference was great enough that it made it difficult for them to quickly and reliably learn to discriminate between pitch and timbral changes. To assure that all of the subjects who participated in the categorization experiment had the same level of competence in discriminating between pitch and timbral changes, an entrance test was created which needed to be successfully completed as a precondition for subject participation in the study. The entrance test was divided into two parts: the first part was a timbre discrimination test and the second part was a pitch discrimination test.

The procedure was identical to the discrimination procedure described in Chapter 2 (Section 2.5.1); with the only difference being that the step size of pitch and timbre variants were chosen as described below.

3.1.1 Subjects

Six subjects (1 male), out of 9, that were able to complete the discrimination task for the N set described in the previous chapter were used to validate the entrance test. In addition, a group of 69 new subjects (23 males) were tested as a part of recruiting efforts for the categorization experiment. Details of the performance of all 75 of these subjects are given below.

3.1.2 Stimuli

The sound stimuli belonged to the N set described in Chapter 2 (Section 2.1.2) and were synthesized as previously described. In both the timbre and the pitch discrimination tasks, four levels were used for the dimension relevant to the discrimination and three levels were used for the irrelevant dimension. The distances between adjacent stimuli along the relevant dimension were one step; the distances along the irrelevant dimension were 1 and 2 steps when the irrelevant dimension was timbre and 1 and 3 steps when the irrelevant dimension was pitch (see Figure 3.2). One step is equivalent to one semitone for pitch and to two spectral steps, according to the formulae 2.1 and 2.2 (Chapter 2, Section 2.1.2), after the adjustment described in Section 2.4, for timbre. The 12 sounds were combined in all possible pairs (yielding a total of 144 pairs of stimuli).

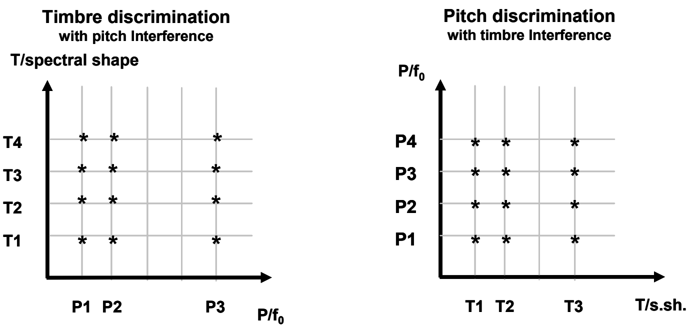


FIGURE 3.2. Entrance tests. Schema of the pitch and timbre levels.

3.1.3 Procedure

A training session and a test session, identical to those described for the discrimination test in Chapter 2, Section 2.5.1, were used for each part of the test. The time course of each trial for the training and test phase (Section 2.5.1) is reported in Figure 3.3.

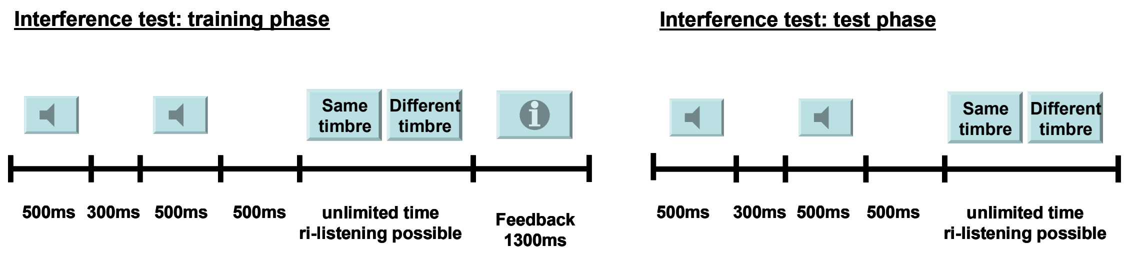


FIGURE 3.3. Entrance tests. Time course of a trial in the training phase (left) and in the test phase (right).

3.1.4 Analysis

Both pitch and timbre discrimination tests were analyzed as follows. The test phase consisted of 3 repetitions of 144 trials. Correct trials were coded as 1 and incorrect trials were coded as 0, thus the average outcome over trials represents discrimination accuracy. Each pair of stimuli had a distance along the relevant dimension of 0, 1, 2, 3, steps, and a distance along the irrelevant dimension of 0, 1, 3, 4 steps when the irrelevant dimension was pitch, and of 0, 1, 2, 3 steps when the irrelevant dimension was timbre. For each single subject the trials were sorted into three groups according to the relevant dimension differences: far (2 or 3 steps), close (one step), and identical (zero steps). Table 3.1 shows the number of each of these comparisons employed in the procedure.

In order to select subjects for the categorization experiment, the following criteria were applied to each single subject in each discrimination task:

1. The mean performance accuracy for the far and for the identical group of stimuli shown in the table (the groups made of stimulus pairs where the sound difference along the relevant dimension was either zero or greater than 1 step) had to be better than 70%.
2. The slopes of the line that fitted response accuracy as a function of irrelevant stimulus dimension distance for far, close and identical stimuli had to be greater than -0.075 for timbre discrimination (first task) and greater than –0.1 for pitch discrimination (second task). These values represent the slopes of a line that goes from perfect accuracy (100%), when the distance along the irrelevant dimension is zero, to 70% when the distance along the irrelevant dimension is maximum.

	$\Delta_{IRREL}=0$	$\Delta_{IRREL}=1$	$\Delta_{IRREL}=3$ (pitch steps) $\Delta_{IRREL}=2$ (timbre steps)	$\Delta_{IRREL}=4$ (pitch steps) $\Delta_{IRREL}=3$ (timbre steps)
$\Delta_{REL}=2,3$	18x3	12x3	12x3	12x3
$\Delta_{REL}=1$	18x3	12x3	12x3	12x3
$\Delta_{REL}=0$	12x3	8x3	8x3	8x3

TABLE 3.1. Stimulus comparisons employed in the entrance tests procedure.

3.1.5 Results

The mean accuracy and the regression line for the subjects accepted and rejected in the two tasks are shown in Figure 3.4.

75 subjects (51 females) entered the timbre discrimination test. 57 (36 females) passed the training phase and completed the timbre test; 21 of these 57 (16 females) passed it and were admitted to the pitch test. No one was rejected during the training phase of the pitch test, but 3 subjects did not pass the standard for minimum performance during the test. Thus 18 subjects (14 females) passed both the timbre and the pitch tests, and were admitted to the categorization experiment. Twelve subjects (7 females), randomly chosen from among those that did not pass the timbre discrimination test, were also given the pitch task. All of them completed the test but only 4 (1 female) had a performance above the criterion.

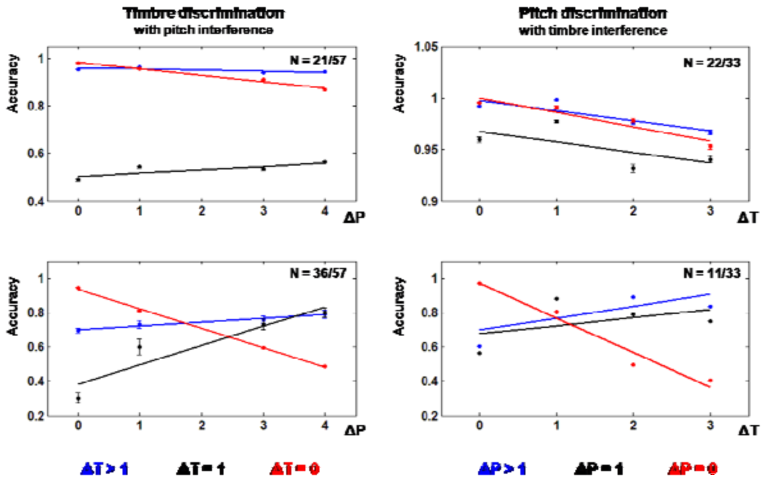


FIGURE 3.4. Summary of the result of the entrance test: mean regression lines for accepted subjects (upper panels) and for rejected subjects (lower panels). Pitch discrimination results include all subjects tested. Error bars report standard errors.

3.2 Categorization experiment

3.2.1 Design of the categories

The sound stimuli used in the categorization task have been described in Chapter 2. They consist of filtered noises with 10 components, and the percepts that they evoke vary along the dimensions of pitch and timbre (from "tonal" to "noiselike").

For category training the sounds were organized into two arbitrary artificial categories, named A and B, on the basis of their pitch information (timbre was not relevant). The two artificial categories were defined in terms of borders (range) and inner structure, as described below

(Figure 3.5). During the test phase, the sound space was sampled in a homogeneous way (Figure 3.5, lower panel, red dots).

Timbre was irrelevant for categorizations. Stimuli could have one of five different timbres, each selected to have a distance of two spectral steps from adjacent stimuli in the series. These different timbres were presented with the same probability throughout the experiment.

Pitch was the relevant dimension for categorization. A range of less than two octaves (from 200 to 682 Hz) was chosen. The basic step was 1.25 semitones. During training, each “level” of the pitch of a stimulus was represented by three separate values of pitch, chosen within a range of a half semitone (Figure 3.5 upper panel).

The two categorical pitch classes were non-overlapping. Class A was defined in the range 215 – 332 Hz, class B was defined in the range 412 – 635 Hz. Each class had an inner structure in which sounds coming from the center of the class were heard during training (both tasks) three times more often than sounds coming from the flanks. Very high and very low pitch levels that do not belong to the classes were also presented, with a probability of one half of that of the flank sounds, and they were explicitly indicated as not belonging to either pitch category. Sounds from the region between the classes were not presented during training.

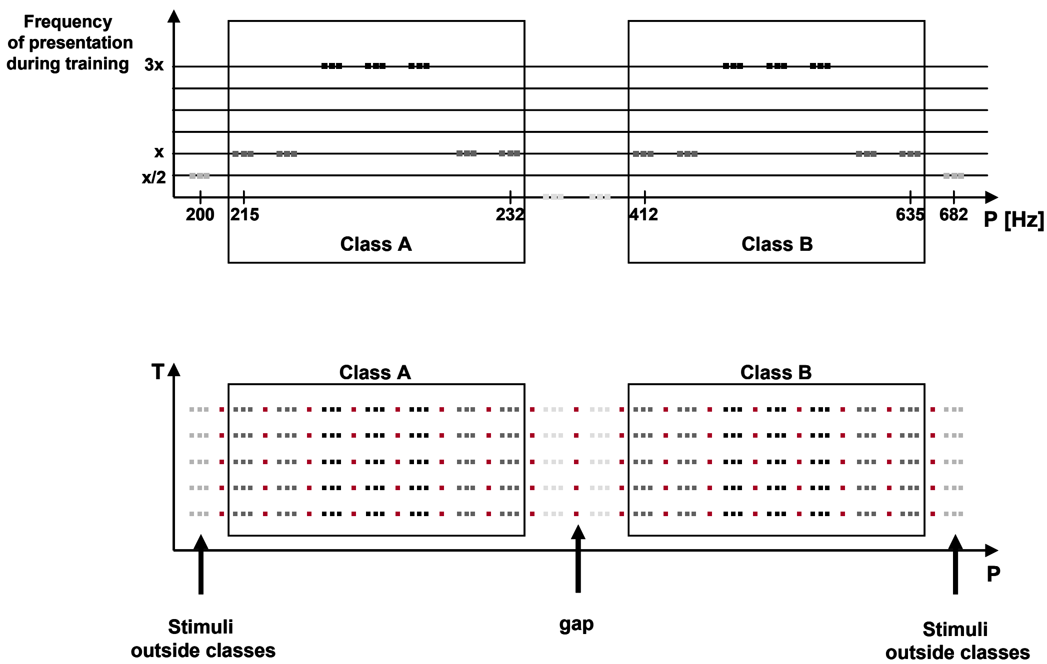


FIGURE 3.5. Lower panel: sampling of pitch and timbre space for the test tasks (red dots) and the training tasks (dots from grey to black). Upper panel: probability of hearing sounds during training as a function of pitch. For timbre levels the probability is uniform.

3.2.2 Format of the Experiment

To study category learning, a pre-test / training / post-test paradigm was adopted.

Before and after training, subjects' abilities to classify sounds according to the artificial categories were tested with an identification task, and the ability to discriminate pairs of sounds along the relevant stimulus continuum was tested behaviorally using a set of auditory staircase procedures. An EEG component called the "mismatch negativity" (MMN, Näätänen, 1992) that is believed to correlate with discrimination ability was also measured before and after training.

Categorization training ensured that the subjects' ability to assign stimuli to the two artificial categories improved to some stable level. Training was composed of two tasks, an identification task in which subjects were asked to sort sounds into the two categories, and a pair-judging task in which subjects were asked to decide whether pairs of sounds belonged to the same or to different categories. The procedures for the training and test tasks are explained below.

3.2.3 General methods

Subjects: 18 subjects (4 males and 14 females, age 18-35 years) were selected as described above in Section 3.1. Of these subjects only 11 (8 females) completed all testing procedures (the others dropped out for various personal reasons at different stages). All subjects gave informed consent for behavioral and EEG experiments, and were paid for their participation. After each session only half of the payment was remitted; the remaining half was paid at the completion of all procedures. The amount of payment depended on the task as follows: subjects were paid 8€ per hour for participation in the behavioral pre- and post-tests and 10 € per hour for participation in the EEG experiment. In each training session (lasting around 1-1.5 hours), to motivate learning, subjects received a reward that was proportional to their best performance over the two training tasks, with the maximum payment being 10 €.

The procedures had different durations for different subjects both in terms of sessions and in terms of time. Two daily sessions were spent for the pre-training and post-training tests, one session for the behavioral tests (lasting 1-1.5 hours) and one for the EEG recordings (lasting from 2.5 to 4 hours, as some subjects were tested in only one condition and others were tested in both conditions). The training sessions (including both tasks and lasting around 1-1.5 hours per day) varied in number between 6 and 10, depending on the outcome of training. The time interval between different training sessions could not be controlled by the experimenter: it was different for different subjects and it was not always constant over time for the same subject. The minimum was one day and the maximum was 15 days. Overall, subjects were involved in the training sessions from a minimum of 7 days to a maximum of 48 days (see table 3.1).

	s1	s2	s4	s5	s7	s11	s12	s13	s16	s18	s19
number of training sessions	7	7	8	8	8	6	10	8	9	6	9
duration of training in days	32	7	26	35	9	35	40	23	48	46	47

TABLE 3.2. Duration of training in number of sessions and days for each subject.

Stimuli: The construction of the sounds has been previously described in Chapter.2, (Sections 2.1.2 and 2.4). All sound stimuli were 300 ms long including 30 ms rise and fall time, except for the sounds used for the EEG experiment, which were 100 ms long including a 10 ms rise and fall time. The levels of pitch and timbre are described for each task below.

Procedure: The behavioral tests took place in a quiet room. Subjects were seated in front of a computer monitor and keyboard. The sounds were amplified and delivered binaurally through Sennheiser HD 580 headphones. Responses were indicated by key presses on the keyboard. Subjects were instructed to choose a comfortable level of sound (which they could adjust by rotating a knob) at the beginning of the experiment and to leave the level the same throughout the entire test series. The EEG recordings took place in a sound-attenuated room. Subjects were seated in a comfortable chair in front of a computer screen at a distance of approximately 1.5 m. Sounds were amplified and delivered binaurally through E-A-RTONE 3A insert earphones (Aearo Corporation, Indianapolis, IN).

3.2.4 Pre-test and Post-test:

Identification task

The stimuli for the identification task sampled the timbre and pitch dimensions homogeneously (Figure 3.5). The timbre levels were the same used in the training phase, and the pitch levels were intermediate to the levels used in training. There were 17 pitch levels spaced 1.25 semitones apart and 5 timbre levels for a total of 85 sounds (Figure 3.5, lower panel, red dot).

The identification session started with 8 examples of pairs of sounds. In each pair the first sound belonged to class A, and the second sound belonged to class B. Timbre was varied randomly. The inter-sound interval within a pair was 3 seconds while the time between each example pair was decided by the subject by key pressing. All sounds were presented singly in random order for identification. Each sound could be classified as “A”, “B” or as “neither A nor B”. Sounds could not be played more than once. There was no time limit for responding and no feedback. Overall there were 2x85 trials, allowing 10 repetitions for each pitch level under test. There were no breaks. On average the test lasted 15-25 minutes.

The time course of each trial was as follows. One sound (300 ms each) was played. After 200 ms, a popup window appeared on the monitor asking the subjects to identify the sound. After the response was made there was an interval of 1000 ms before the next trial started (Figure 3.6).

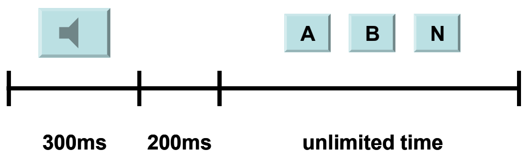


FIGURE 3.6 Time course of one trial in the identification test.

Auditory staircase task

The rational for this test is to investigate the perceptual discrimination ability of subjects along the pitch continuum. Three pitch levels were selected as central stimuli for the staircases: 258 Hz (from the centre of class A), 369 Hz (from the untrained region), and 530 Hz (from the centre of class B). Six staircases, one on either side of the central frequencies, were run in a random order (Figure 3.7 upper panel). All sound stimuli in the staircases had the same intermediate timbre level (the third level over five in the series from tonal to noiselike).

The experimental paradigm was a weighted 2-down 1-up adaptive procedure (Levitt, 1971; Kaernbach, 1991). On each trial, a pair of stimuli was presented. The task of the subject was to indicate which of the two stimuli was higher in pitch. Subjects were not allowed to have a trial repeated. They had unlimited time to answer and received no feedback on the correctness of the response (Figure 3.7 lower panel). Of the stimuli presented in a trial, one (position randomized) had f_0 equal to the center frequency under test, and the other had a frequency of $f_0 \cdot 2^{n/480}$, where n was varied adaptively in steps of one unit, leading to a variation of f_0 in steps of a fortieth of semitone. The initial difference was 2 semitones, which was also the maximum difference possible. When two correct answers were given in a row, the parameter n was decreased by half, on a wrong answer n was increased by 1.5 times. This leads to asymptotic performance at the point of 70.7% correct answers (appendix 1; see Levitt 1971; Kaernbach 1991 and Garcia-Perez 1998, for a critical discussion of such procedures).

The staircase continued until 15 reversals or a maximum of 100 trials were reached, and the point of 70.7% accuracy was then determined as the average of all differences from the third to the last reversal. On average the test lasted between 30 and 40 minutes.

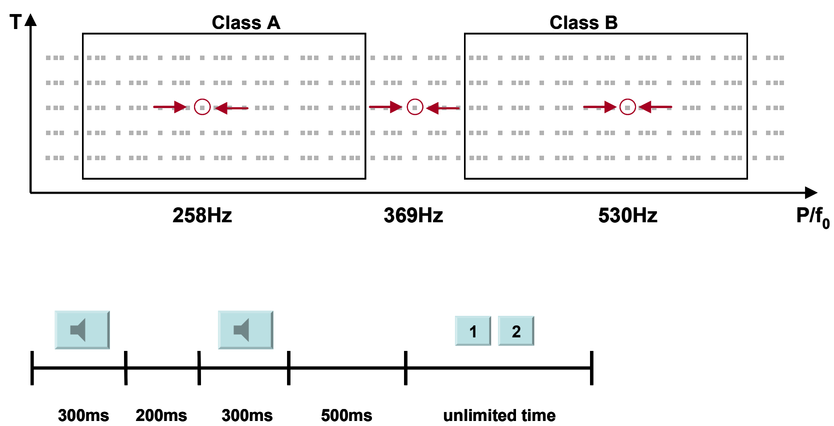


FIGURE 3.7. Auditory staircase tests. Indication of the central frequencies (upper figure) and time course of one trial (lower figure).

3.2.5 EEG recordings:

Procedure: The EEG was recorded in an oddball paradigm. Subjects were seated comfortably in an armchair contained inside a sound-attenuated room and asked to relax and minimize movements and eyeblinks. They were asked to focus their attention on a silent movie (sound off) of their own choice played on a screen and ignore the sounds delivered through the insert earphones (stimulation was binaural and at the same intensity level each subject had chosen for the behavioral test). The sounds used in the oddball procedure are shown in Figure 3.8. The standard was chosen from inside one category and presented 80% of the time; the four deviants, at a distance of 2.5 and 5 semitones on either sides of the standard, were each presented 5% of the time. The standard was always chosen inside one of the classes; the deviants were chosen according to two different conditions. In condition 1, two of the four deviants were lower than the standard and belonged to the central region of the same category as the standard; in condition two, these two deviants were higher than the standard. The remaining two deviants on the other side of the standard did not belong to the same category as the standard: one was chosen in the untrained region between the two categories, and the other was chosen in the flank region of the category that the standard did not belong to (Figure 3.8).

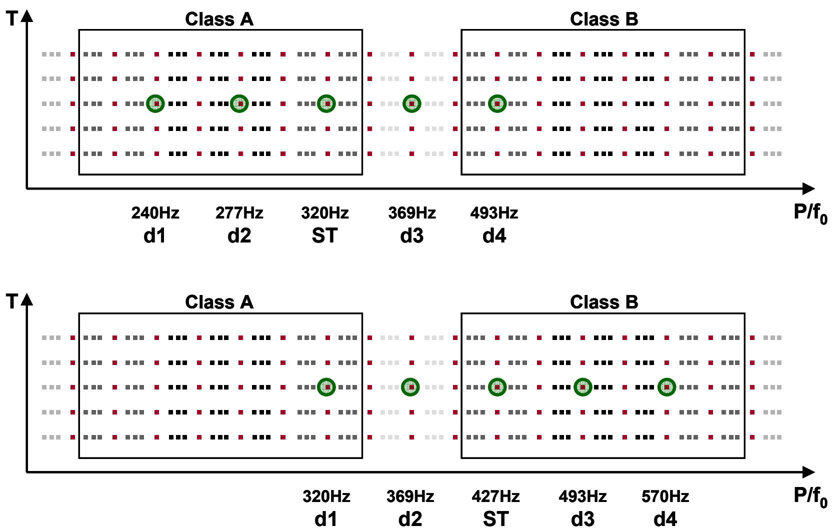


FIGURE 3.8. Schema indicating the position of the oddball sounds in the timbre pitch space.

Each recording session consisted of 5040 sounds (4040 repetitions of the standard and 250 of each deviant). Sound order was pseudo-randomized by imposing a rule such that each deviant had to be preceded by at least three standards. The oddball sequence was divided into two blocks to allow subjects a small break of 5 minutes after the first 2500 stimuli (~35 minutes of recording). During the break the subjects could drink and talk, but none of them got up from the recording armchair. Twenty repetitions of the standard were added at the beginning of each recording block to facilitate the formation of the MMN. Each sound lasted 100 ms (included 10 ms of cosine onset and decay) and the stimulus onset asynchrony was 800 ms.

All subjects were randomly assigned to one condition and a few subjects were tested in both conditions: 4 subjects (3 females) in the first condition, 4 subjects (2 females) in the second condition, and 3 subjects (3 females) in both. For the subjects tested in both conditions, the order was random and counterbalanced across subjects. The two recording sessions were made on the same day with 30 minutes break between them.

Recordings: EEG was continuously recorded using an ActiveTwo data acquisition system from BioSemi (www.biosemi.com) with 128 scalp sensors, placed according to the BioSemi coordinates (see Figure 3.9). Additionally, electrical activity was recorded at both mastoids and at the tip of the nose for offline referencing. Electro-ocular activity was recorded from two bipolar montage electrodes. The vertical EOG was recorded from the left eye by a supra- and infra-orbital electrode (the supra-orbital electrode was a frontal one from the electrode-cap) and the horizontal EOG was recorded from the outer canthi of both eyes. Recordings were reference-free. An additional channel was used to record the sound stimuli. All channels were low pass filtered online with a cutoff frequency of 417 Hz and digitized with 24-bit resolution at a sampling rate of 2048 Hz before being stored in a computer.

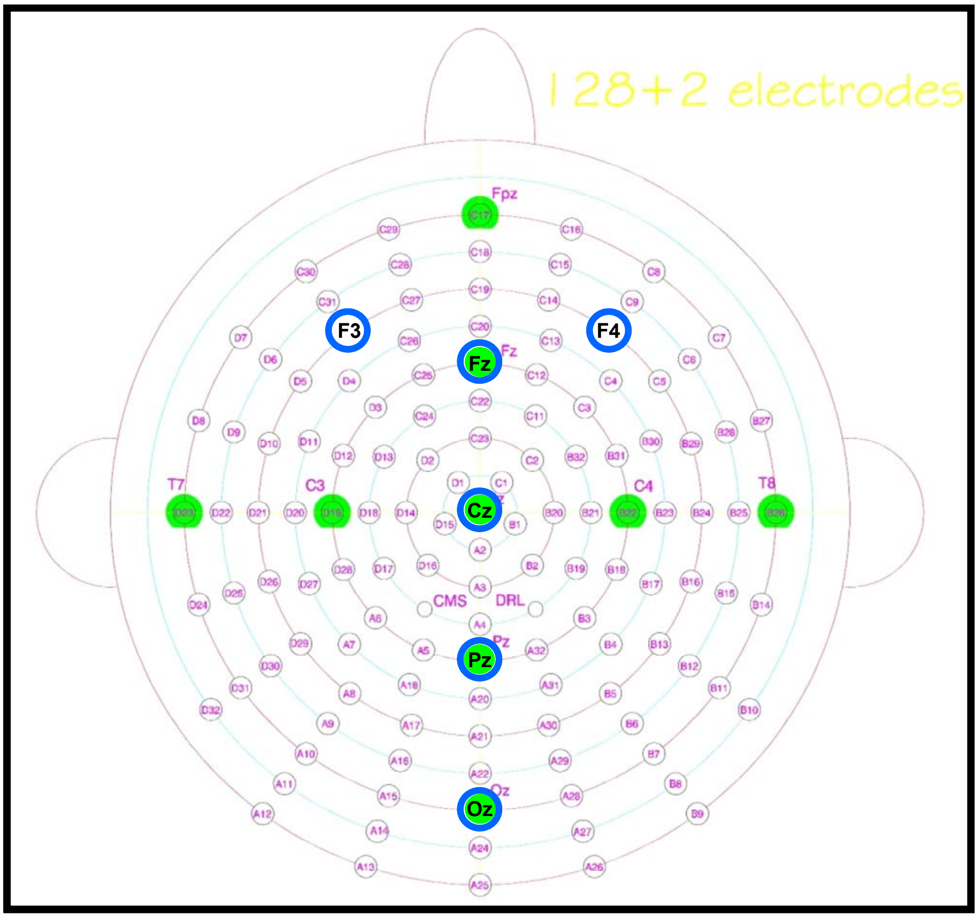


FIGURE 3.9. Schema of the electrode positions of the standard BioSemi 128-channel headcap. The positions corresponding to the 10-20 system sites are in green. The blue circles indicate the sites used in the analysis

MMN extraction and quantification: EEG data were down-sampled to 256 Hz and analyzed with the MATLAB toolbox EEGlab (Delorme & Makeig, 2004). The time marks for the onset of each sound were extracted from the sound channel and stored. The analyses were conducted on a sub-sample of 6 scalp sites: Fz, Cz, Pz, Oz, F3, F4. The coordinates of Fz, Cz, Pz, Oz were precise, but those of F3 and F4 were only approximated since there is no correspondence between BioSemi coordinates and the coordinates of the extended International 10-20 system for these sites. The EEG recordings were referenced off-line to the average mastoid in order to maximize the MMN amplitude at frontal sites (Kujala, Tervaniemi and Schröger, 2007).

The channels were bandpass filtered with a finite impulse response filter (FIR) with the following specifications: 768 points high pass filtering, cut off frequency 1 Hz, transition band width 0.2 Hz; 24 points low pass filtering, cut off frequency 30 Hz, transition band width 5 Hz. Based on the stimuli time marks, epochs of 800 ms length, including 200 ms pre-stimulus baseline, were cut. The ERPs to the first ten standards in each block and to every standard following a deviant were removed to ensure that each ERP under analysis would be preceded by an ERP to a standard stimulus. After baseline subtraction, each epoch showing a voltage exceeding ± 50 µV on one of the channels (including the bipolar HEOG and VEOG) was rejected. On average, each session provided 2066 trials for the standard and around 170 for each deviant.

For each participant, the average response was computed separately for all conditions and the difference waves were obtained by subtracting the mean response to the standard from the mean response to each deviant. Grand averages were also computed for the group of subjects as shown in Figure 3.10 for the pre-test recordings.

The MMN was quantified at the Fz site, for each single subject and for each deviant. It was defined as the negative deflection in a fixed time window and its amplitude was quantified as the average value during this time window. The time window was defined as the interval between 70 ms and 200 ms after stimulus onset, and was chosen to be broad in order to include all subjects' latency variations. Note that this definition includes components of an evoked potential called the N1 (Näätänen & Picton, 1987) in the measure of MMN; in practice, it is very hard to make an objective separation between these two components in the kinds of recordings collected here.

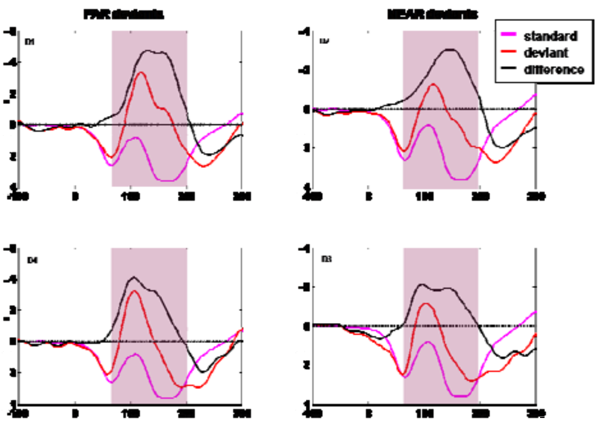


FIGURE 3.10. Grand average of the standard and deviants across all subjects in the pre-training EEG session. The site is Fz. The light red shadows indicate the time window in which each individual MMN is measured as described in the text.

3.2.6 Training:

Each training session included the identification task, always performed first, and the pair-judging task.

A number of initial training sessions were run with no time limits in each trial, to help subjects learn the categories. When the performance of the subject was considered to be good and stable according to the predefined criteria given below, the same two tasks were modified to include a time limit for responses, in order to help make the categorization process automatic. The criterion was that performance accuracy for both training tasks had to be no less than 85% correct for three sessions over four in a row, or no less than 80% correct for four sessions over five in a row. A second, similar criterion was adopted to terminate training: successful subjects performed with accuracy higher than 85% for three sessions over four in a row in both time-limited tasks were passed on to the testing phase. Training could also be ended by completing a maximum of 5 time-limited sessions.

Identification task

The identification training task was the same as the identification test task with two major differences: first, the sound stimuli that were presented to the subjects came from the training set and reflected the presentation probabilities described in Section 3.2.1 (see Figure 3.5); second, subjects were given feedback after each trial.

In the un-limited time version, the feedback was presented after each wrong trial in the form of a written message that provided the correct answer. Subjects moved to the next trial by closing the feedback window. On a correct trial there was no feedback and the subjects moved on to the next trial after 1 second delay (Figure 3.11). There were 270 trials presented randomly; 5 trials for each of the extreme pitch levels, 10 for each level from the categories' flanks and 30 for each level from the categories' centre (see Figure 3.5 and 3.12). For each pitch level there were three pitch values covering a region of half of a semitone and five timbre levels. Thus only the stimuli from the centre of the categories were repeated twice in a session. Every 90 trials, subjects were informed about their performance and encouraged to do better; they could have a break, and had to press a key to continue the task. Only the first of the unlimited time identification experiments was preceded by an example of a sound from class A and an example of a sound from class B. All the subsequent sessions had no examples.

In the time limited version, the procedure was very similar, but subjects were given 900 ms to identify the sound. If they succeed in inserting an answer they were given fast feedback (800 ms) in the form of a colored screen (red for wrong and green for right); if they were too slow in inserting the answer the feedback was a yellow screen (800 ms). The reaction times were recorded. Subjects were allowed one break in the middle of the test (after 135 trials) and were informed of their score only at the end of the task.

The time course of each trial for the training identification test in the time-unlimited and the time-limited versions are shown in Figure 3.11

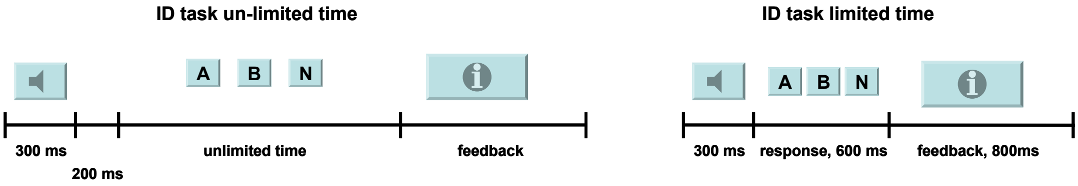


FIGURE 3.11. Identification training task. Time course of one trial in the unlimited time paradigm on the left and limited time paradigm on the right.

Pair judging task

The sound stimuli used in this test came from the training set. They were arranged in pairs. One sound always belonged to one of the categories; the other sound could be internal to the same category, it could belong to the other category or to one of the extreme regions outside both categories. For each pair the participants had to decide whether the two sounds belonged to the same category or not – they did not have to indicate which category the stimuli belonged to.

The probability of presentation for each sound reflected the inner structure of the categories. In each session there were 270 pairs presented in random order, 140 from different categories (120 crossing the border between A and B and 10 crossing each of the two extreme borders) and 130 from inside one category (65 for each category). Pairs were made anew for each session. Figure 3.12 shows the frequency of presentation (upper half of the figure) and the schema for the construction of stimulus pairs.

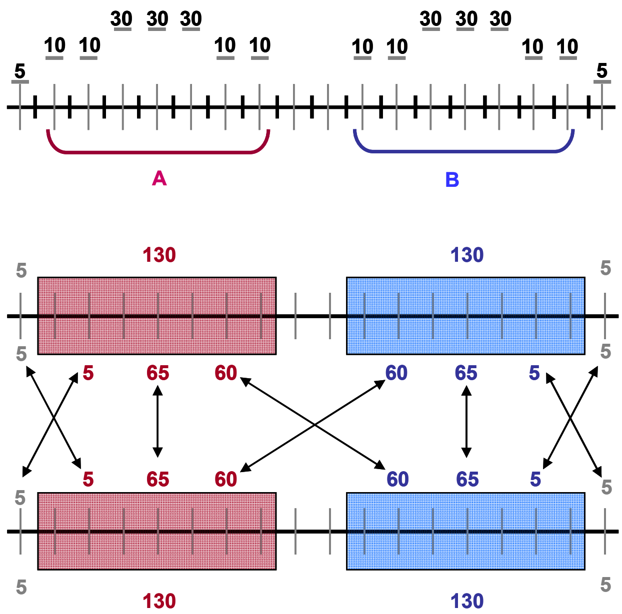


FIGURE 3.12. Pair-judging task: frequency of presentation of each pitch level (upper panel) and schema of the construction of pairs (lower panel).

Before starting with the first training session, subjects were given two examples, an internal pair and a cross-boundary pair. Feedback on each single trial and scores were provided in the same way as in the identification training task.

The time course of each trial for the training of the pair- judging tasks in the time-unlimited and the time-limited versions are shown in Figure 3.13. The two sounds in a pair (300 ms each) were played with an interval of 2500 ms of silence between them. In the unlimited time version of the task, a popup window appeared after 200 ms asking "Do the sounds belong to the same category or to different categories?" and there was no time limit for answering. In the time-limited task, the question was not repeated and subjects were only allowed to insert the answer (same or different) from the onset of the second sound to 900 ms after. After a response was made, a feedback message appeared for 800 ms. In both versions of the task, there was an interval of 1000 ms before the next trial started.

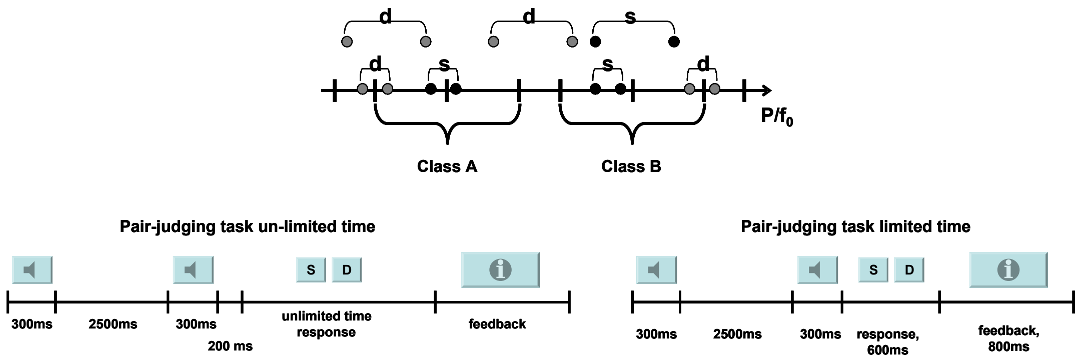


FIGURE 3.13. Pairs-judging task. Schema of the construction of the pairs (upper figure) and time course of a trial in the un-limited time (lower left figure) and limited time (lower right figure) paradigms.

Chapter 4

In Chapter 3 the procedures employed for selection, training and testing of the participants for the categorization experiment were reviewed. The results of the experiment are analyzed and discussed here. Since the goal of the experiment was to investigate the effect of category learning on perceptual discrimination abilities, the results of the training and identification tasks were used to sort the subjects into homogenous groups of learners based on their performance. The discrimination abilities of these groups of learners were then compared. Categorical perception theory predicts that compressions and/or expansions of perceptual distances arise as a consequence of category learning, which yields specific predictions about changes in the discrimination abilities of different groups of subjects for the different sounds presented in the experiments.

4.1 Training outcomes

To conclude training each subject had to complete the last three trials with an accuracy of no less than 85%. All subjects fulfilled the requirements with the exception of one subject whose training stopped after the fourth time-limited session (S2). The accuracy of this subject in the identification task was over the 85% threshold, and performance in the pair-judging task increased monotonically from 60% in the first session to 80% in the fourth, so the subject was considered successful and included in the analysis. Table 4.1 reports the number of training sessions for each subject.

	s1	S2	s4	s5	s7	s11	s12	s13	s16	s18	s19
number of UNLIMITED TIME training sessions	3	3	3	4	4	3	4	3	4	3	4
number of LIMITED TIME training sessions	4	4	5	4	4	3	6	5	5	3	5

TABLE 4.1. Number of training sessions for each subject.

Data from the identification and pair-judging tasks and for the time-limited and the time-unlimited phases were analyzed separately. To confirm learning, the subjects' mean accuracies and reaction times at the first and last sessions were compared, for each task and each training phase. The last session could be the third, fourth, fifth or sixth depending on the subject and the phase of training. In the unlimited time training phase, due to a technical failure, the accuracy of subjects 12 and 16 in the first session of the identification task -unlimited time phase- was not stored, thus the second session was used for the analysis. The comparisons are summarized in Table 4.2.

	s1	S2	s4	s5	s7	S11	s12	s13	s16	s18	S19
UNLIMITED TIME comparison between training sessions	1-3	1-3	1-3	1-4	1-4	1-3	2-4	1-3	2-4	1-3	1-4
LIMITED TIME comparison between training sessions	1-4	1-4	1-5	1-4	1-4	1-3	1-6	1-5	1-5	1-3	1-5

TABLE.4.2. For each subject the results in terms of accuracy and reaction times were compared for the sessions indicated in the table.

Normality of these data were assessed using a Lilliefors test, and homogeneity of variances were assessed using a Bartlett test. Because of departures from normality, the accuracy data were arcsine transformed and the reaction time data were submitted to a logarithmic transformation (base e). One tailed t-tests were performed on the transformed data rather than two-tailed tests because of the pre-existing directional prediction that accuracy and reaction time would both improve (greater accuracies and shorter reaction times) with training. The figures depict the un-transformed data.

4.1.1 Identification training

The performance of each subject was monitored during training by looking at the average accuracy over all trials in each session. After training, accuracies and reaction times of each pitch level were averaged, so that the performance for the sounds close to the category borders and the central sounds were equally represented. Figure 4.1 displays the mean accuracy of each subject (plotted as a function of training sessions) separately for the two training phases (unlimited time and limited time). The average reaction times for each subject are also shown for the time limited phase of training.

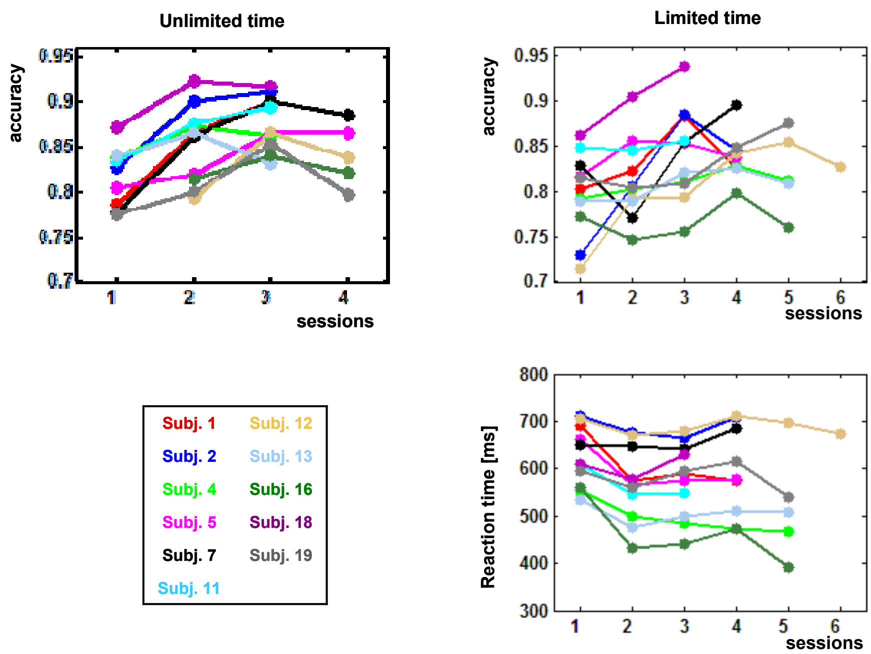


FIGURE 4.1. Results for the single subjects in the identification training task. Top left: accuracy in the unlimited time phase. Top right: accuracy in the limited time phase, Bottom right: reaction times in the limited-time phase. Each subject is represented by a different color.

The mean accuracy significantly increased in both phases, while the mean reaction time significantly decreased (unlimited-time phase: df = 10, t = -4.2591, p = 0.0008; limited-time phase accuracy: df = 10, t = -3.7259, p = 0.0020; limited-time phase reaction time: df = 10, t = -2.8108, p = 0.0092). See Figure 4.2.

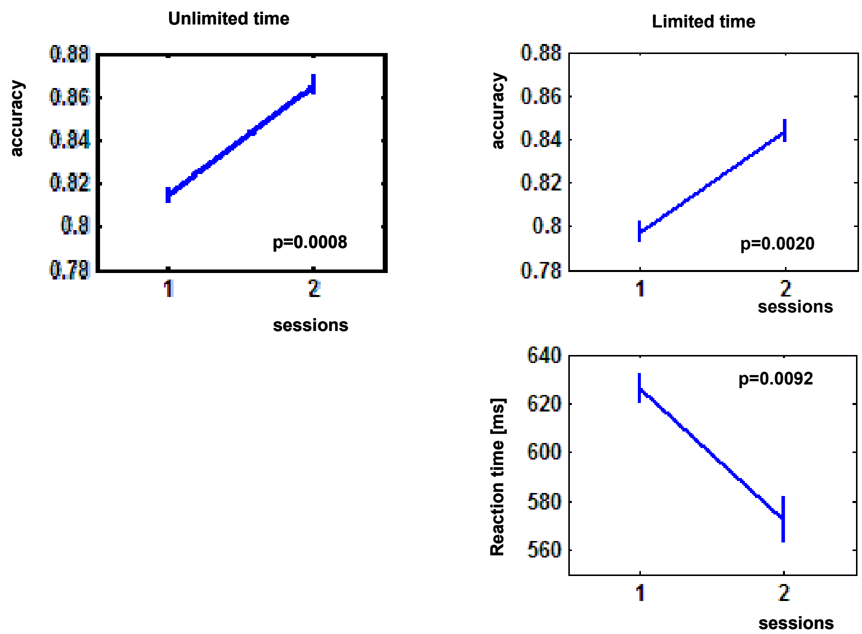


FIGURE 4.2. Identification training task. Average accuracy across subjects for the unlimited (top left) and limited time phase (top right). Average reaction times across subjects for the limited time phase (bottom right). Error bars indicate one standard error.

4.1.2 Pair-judging training

In Figure 4.3 the mean accuracy of each subject is plotted as a function of training sessions, separately for the two training phases (unlimited time and limited time). The average reaction times for each subject are also shown for the time limited phase of training.

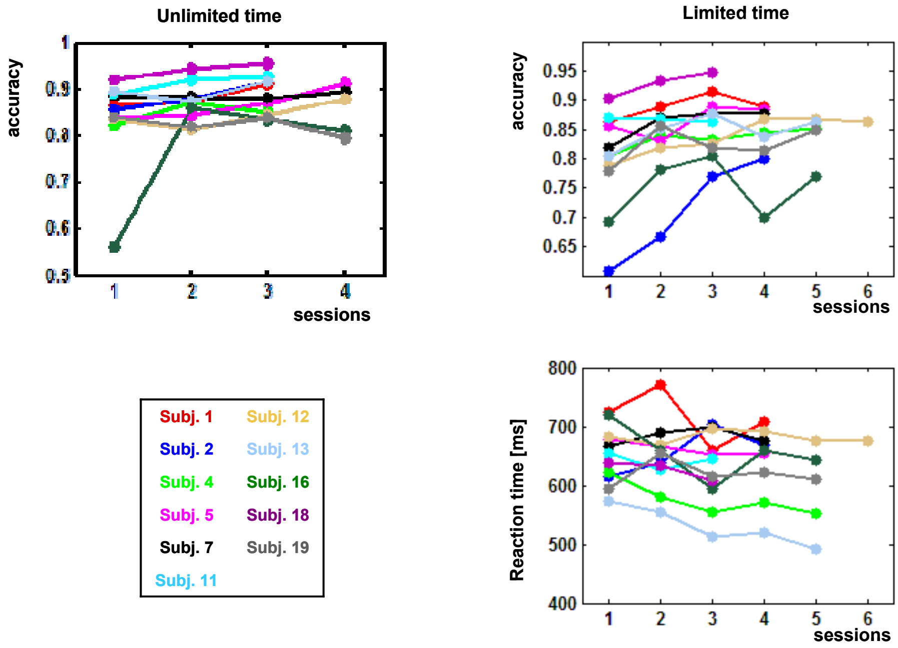


FIGURE 4.3. Results for the single subjects in the pair-judging training task. Top left: accuracy in the unlimited time phase. Top right: accuracy in the limited time phase, Bottom right: reaction times in the limited-time phase. Each subject is represented by a different color.

As shown in Figure 4.4, mean accuracy significantly increased in both phases, while the reaction time decreased almost significantly (unlimited-time accuracy: df = 10, t = -3.0325, p = 0.0063; limited-time accuracy: df = 10, t = -5.0037, p = 0.0003; limited-time reaction time: df = 10, t = -1.6819, p = 0.0618). The results for the first phase of training remain significant when the accuracy data from subject 16 (whose performance in the first session was particularly low –see Figure 4.3) were removed from the analysis (df = 9, t = -3.4011, p = 0.0039). It is notable that with the exception of subject 16, most subjects were already performing above chance in their first session in both phases, suggesting that the task was not extremely difficult.

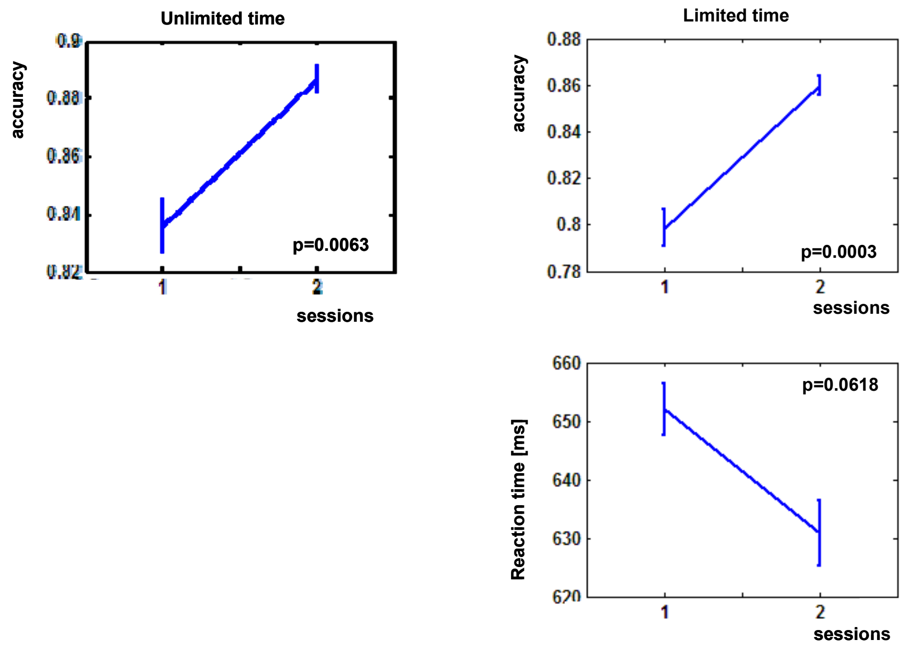


FIGURE 4.4. Pair-judging training task. Average accuracy across subjects for the unlimited (top left) and limited time phase (top right). Average reaction times across subjects for the limited time phase (bottom right). Error bars indicate standard error.

4.2 Comparison of pre- and post-training

The performances of the subjects in the identification and the auditory staircase tasks were compared before and after training.

4.2.1 Identification task

Figure 4.5 shows the identification curves, averaged across subjects, as a function of pitch levels and timbre levels. Improvement in identification performance was assessed at a group level and for each single subject.

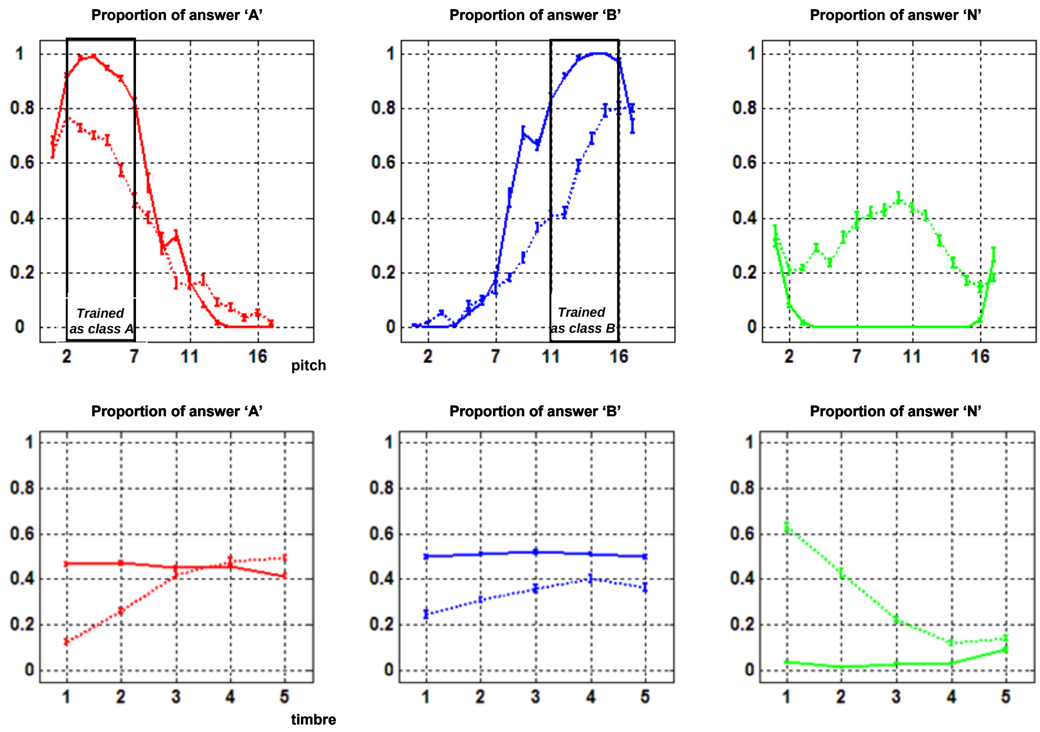


FIGURE 4.5. Grand average identification curves before (dotted lines) and after training (continuous line) as a function of the pitch levels (upper plots) and of the timbre levels (lower plots). Sounds from pitch level 2 to pitch level 7 are defined as category A during training and sounds from pitch level 11 to pitch level 16 are defined as category B (shown in the first two upper pictures).

At a group level, the extent to which timbre identification was changed by training was analyzed by comparing the slope (β) of the linear regression lines for the curves before and after training. The results showed that training changed the way timbre was perceived (for category A: β_{pre} = 0.0952, β_{post} = -0.0120, F=39.53, p=0.0001; for category B: β_{pre} = 0.0329, β_{post} = -0.0003, F=3.67, p=0.058; for elements outside categories: β_{pre} = -0.01281, β_{post} = -0.0123, F=69.32, p<0.0001), and that timbre was irrelevant to classification after training, except for a trend among the sounds classified as not belonging to any category to include more sharper timbre levels than duller timbre levels. (for category A: df = 53, t = –1.4262, p = 0.1597; for category B: df = 53, t = –0.0350, p = 0.9722; for elements outside categories: df = 53, t = 3.0176, p = 0.0039) Category identification accuracy based on pitch information before and after training was compared for the sounds that were within the category boundaries. Subjects showed significantly improved performance for both category A and B sounds (for class A: N = 11, t = –5.324, p = 0.00034; for class B: N=11, t = –7.2664, p = 0.00003). See Figures 4.5 and 4.7.

At the single subject level, all subjects learned that timbre was irrelevant for classification (all had slopes lower than 0.044, Figure 4.7). All subjects also had accuracy greater than 80 % for each sound belonging to the two categories (see Figures 4.6 and 4.7). This 80% criterion reflects a 50-50 chance of answering A or B in the identification test (since each sound was repeated for identification ten times, the probability of obtaining 8 or more correct answers was 0.0547 and the probability of obtaining 9 or more correct answers was 0.0107, according to the binomial distribution). The assumption of a 50% chance level for the responses A and B is warranted by the very low probability of hearing a sound coming from outside any of the categories during training.

Five subjects (S 1, 11, 12, 13, 18) correctly classified all sounds in the two categories, while the other subjects failed to correctly classify some of the peripheral sounds in one of the categories. For category A, subjects 4 and 7 did not correctly classify pitch level 7, and subject 2 did not correctly classify pitch levels 6 and 7. For category B, subjects 5 and 16 did not correctly classify pitch level 11, and subject 19 did not correctly classify pitch levels 11 and 12.

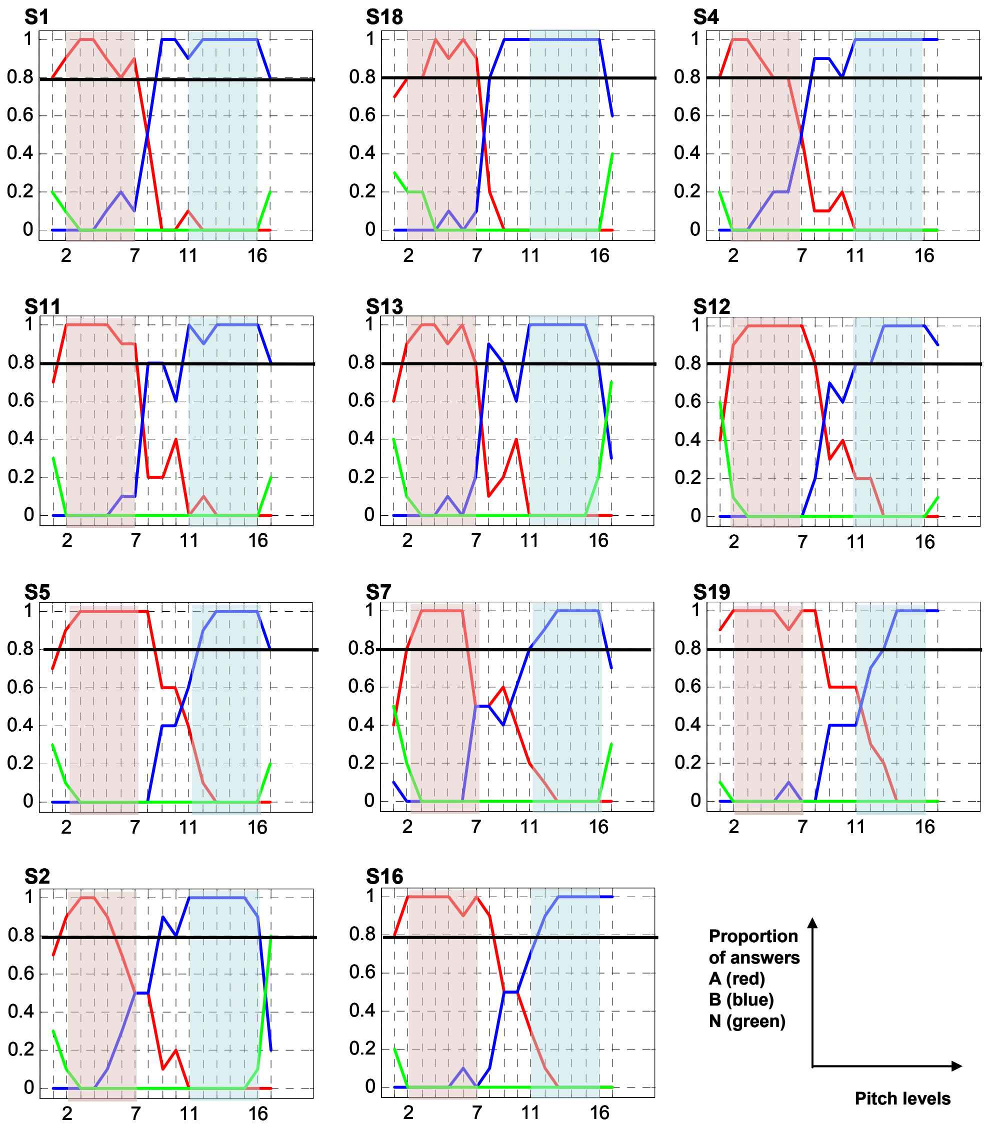


FIGURE 4.6: Subjective identification curves as a function of pitch levels after training. The colored bands indicate the regions that were trained as category A (transparent red) and category B (transparent blue).

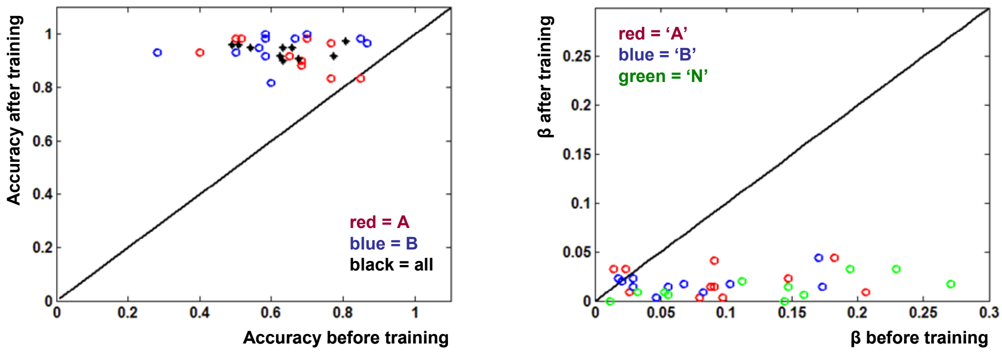


FIGURE 4.7 Scatter plots of accuracy before and after training (left) and the slopes of the regression line between identification response and timbre levels (right). Each data point represents one subject.

Appendix 2.1 presents the identification curves as a function of pitch and timbre before and after training for each subject.

Summary of the results of the identification test:

1. After training all subjects learned that that timbre was not relevant, and classified according to pitch only;
2. After training all subjects were very accurate in classifying items in the middle of the categories (the good members of the categories);
3. Most subjects placed a clear boundary between class A and B after training, but its position and steepness varied;
4. Some subjects extended the classification to higher and lower pitch levels outside the training categories;
5. The performance for the sounds from the periphery close to the border between the two categories was variable across subjects. All subjects stretched one or both categories over the untrained region in between the trained classes.

4.2.2 Auditory staircase task

Six auditory staircases were run around three central frequencies situated in the central region of class A (258Hz), in the central region of class B (530Hz) and in the middle of the gap between the two categories (369Hz).

During training subjects were trained for several sessions to classify sounds very close to the higher and lower staircase central sounds (258 and 530Hz); they were not exposed to frequencies in the region around the median central frequency (369Hz).

The six staircases will be indicated with a capital letter and a subscript. The capital letter specifies the position of the central frequency relative to the trained categories: A indicates the two staircases around 258Hz, as this level belongs to category A, B indicates the two staircases around 530Hz, as this level belongs to category B, M indicates the two staircases around 369Hz, as this level belongs to the untrained region in between the categories. The subscripts ‘up’ or ‘down’ specify whether the comparison sounds being tested in the staircases were higher (‘up’) or lower (‘down’) in frequency than the central sounds..

In the post-training identification task, all subjects were 100% accurate in the classification of the central frequencies of categories A and B, with the exception of subject 4 who had an accuracy of 90% in the identification of the sounds coming from class A.

For each subject, the point of 70.7% accuracy was determined by calculating the mean of all peaks and valleys from the third to the last reversal in each staircase (see Figure 4.8). For each of the staircases an index of change was calculated by taking the difference between the 70.7% accuracy point (AP) before and after training, normalized by the sum of the two accuracy points:.

Change Index = $[AP_{PRE} - AP_{POST}] / [AP_{PRE} + AP_{POST}]$

The change indices in staircases A, B and M were compared separately for the ‘up’ and ‘down’ measures for all subjects (Figure 4.9). Kruskal-Wallis tests revealed no significant differences in the changes in discrimination abilities for the ‘down’ (N=11, df=2, χ^2=1.69, p=0.430) and ‘up’ (N=11, df=2, χ^2=1.69, p=0.429) staircases.

The change indices of all six staircases were not significantly different from zero, as assessed by Wilcoxon Sign Rank Tests (staircase A_{up}: signrank = 17, p = 0.175; staircase A_{down}: signrank = 18, p = 0.206; staircase M_{up}: signrank = 33, p = 1; staircase M_{down}: signrank = 28, p = 0.7; staircase B_{up}: signrank = 32, p = 0.966; staircase B_{down}: signrank = 25, p = 0.519).

These results indicate that the categorization training, although being effective in teaching the subjects to classify the sounds, had no effect on subject’s finer perceptual abilities, as measured by auditory staircases, in the central regions of the categories and in the region between categories.

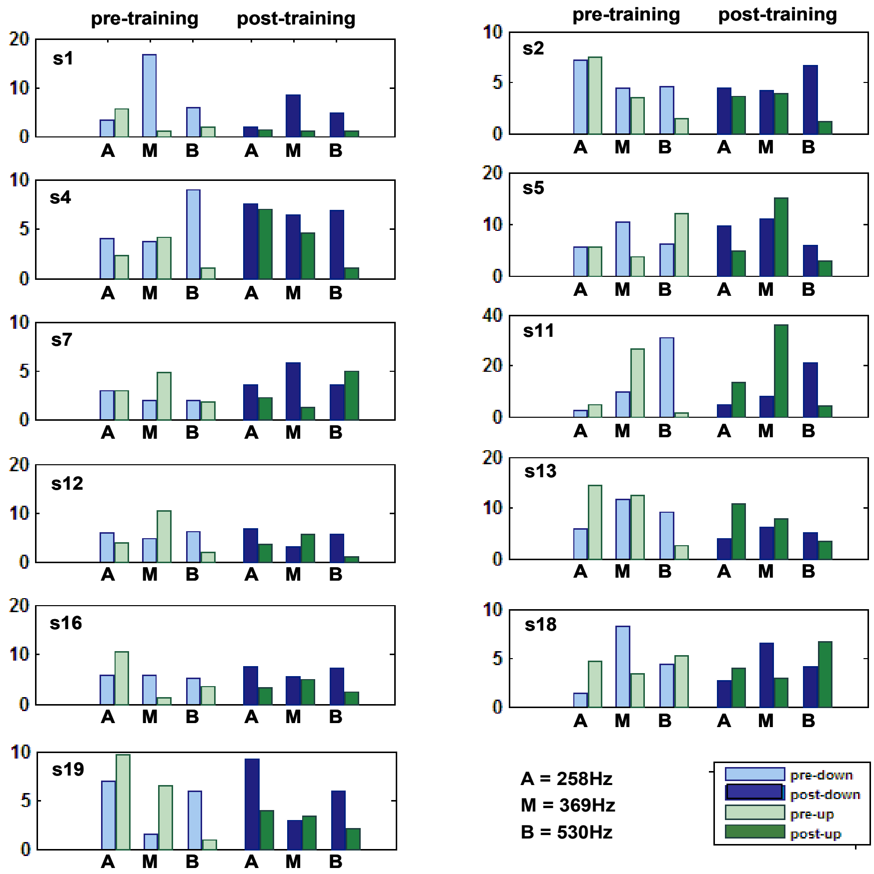


FIGURE 4.8. Results of all staircases for each subject. The point of 70.7% accuracy is expressed in fortieths of a semitone. For each subject, before training results are on the left in light colors; after training results are on the right in dark colors.

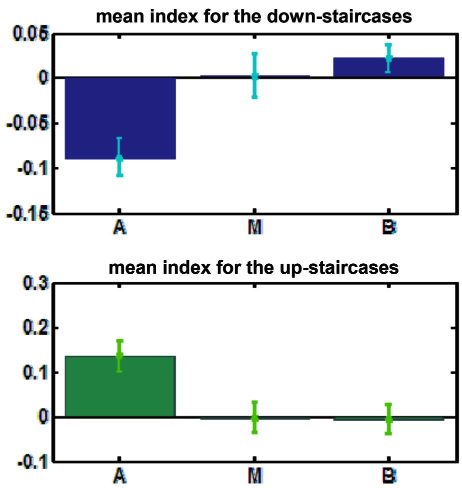


FIGURE 4.9. Mean change index values across subjects and standard errors for the up and down staircases around the three central frequencies A: f_0=258Hz, B: f_0=369Hz, M: f_0=530Hz.

The individual staircases for each subject are presented in Appendix 2.2.

4.3 Analysis and results of the MMN recordings

The goal of the analysis was to assess whether learning to categorize sounds affects the MMN, used as an index of perceptual distance between sounds.

By construction, the oddball sounds were chosen close to the sounds that the subjects were trained to classify as either A or B, with the exception of one deviant coming from the gap between the categories (Figure 4.10).

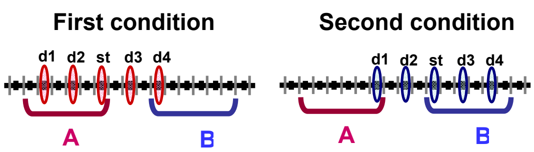


FIGURE 4.10. Schema of the oddball sounds relative to the arbitrary categories.

However, each single subject had their own idiosyncratic borders for the categories after training. Thus, the MMN for the different deviants were analyzed with respect to the subject's classification of the standard and the deviant.

4.3.1 Definition of an MMN amplitude index

To measure the variations in MMN amplitude with time, an **index** was constructed. For each subject and each deviant, the index is given by the difference between the MMN amplitude after and before training, normalized by the sum of the two amplitudes.

Normalized Index = $[A_{MMN}\text{ (after)} - A_{MMN}\text{ (before)}] / [A_{MMN}\text{ (after)} + A_{MMN}\text{ (before)}]$

In this formula each amplitude is taken as the absolute value of the average MMN voltage. Thus, a positive index means that the amplitude of the MMN has increased after training, while a negative index means that the MMN has decreased after training.

Prior to describing the analysis, the behavioral grouping of the subjects will be reviewed.

4.3.2 Behavioral classification

The subjects were grouped according to their performance in the post-training identification test.

Each item in the identification test could be identified as belonging to class A or B or as being external to the two categories. Thus three answers were possible: A, B, N (N meaning none of the classes). During training the subjects were taught that most of the sounds came from A or B and very few from outside the two classes: in any session 48.15% of sounds belonged to class A, 48.15% of sounds belonged to class B and only 3.7% of sounds did not belong to any category.

In the post-training test each sound was repeated for identification ten times. If a 50-50 chance of answering A or B is assumed (based on the low presentation frequency of N sounds), then the probability of obtaining 8 or more correct answers is 0.0547 and the probability of obtaining 9 or more correct answers is 0.0107, according to the binomial distribution.

Following this reasoning a threshold of 0.8 accuracy was set: items that were identified as members of one class 80% of the time or more were considered to belong to that class, while sounds that were identified with lower proportions were not considered to have been classified.

For each subject, the threshold was applied to the identification results of the oddball sounds in order to define their position relative to the subjective category borders. The results are shown in Table 4.3. The entries of the table indicate the subject's classification of the five oddball sounds (in ascending pitch order: d1-d2-st-d3-d4).

SUBJECT	**CLASSIFICATION THRESHOLD = 0.8**	
	1st condition st trained in class A	**2nd condition st trained in class B**
1	**AAABB**	
5	**AAANN**	
13	**AAABB**	
19	**AAANN**	
4		**NBBBB**
11		**ABBBB**
12		**ANBBB**
16		**ANNBB**
2	**AANBB**	**NBBBB**
7	**AANNB**	**NNBBB**
18	**AAABB**	**ABBBB**

TABLE 4.3. Behavioral grouping according to the classification of the oddball sounds in the post- training identification test; st = standard.

Sounds d1 and d2 and sounds d3 and d4 have the same distance from the standard, (2.5 and 5 semitones, respectively). Sounds d1 and d2 were called "close deviants" and sounds d3 and d4 were called "far deviants". Close and far deviants were analyzed separately.

Relative to the behavioral classification three groups were made (Table 4.4):

- Two-class: the standard and one deviant were classified as belonging to one category, while the symmetric deviant was placed in the other category (AAB or ABB);
- One-class: the standard and one deviant were classified as belonging to the same category, while the symmetric deviant was not clearly classified, thus being placed in between categories (AAN or NBB);

- No-class: the classification did not reflect the categorical training; either the standard was not classified in a category, or the border between the two categories was shifted so much that all three items were classified in the same category (ANB, ANN, NNB, BBB).

CP theory makes specific predictions for the first two behavioral groups, stating that, after category learning, pairs of sounds that are classified together would become perceptually more similar, and pairs of sounds straddling the border of a category would became perceptually more different. Thus, the first two behavioral groups were analyzed. Because some subjects were tested in both the 1st and 2nd oddball conditions, prior to the analysis a choice was made about which condition to include in the statistical analysis. The rationale for the choice was to have the behavioral groups for this analysis as homogeneous as possible.

Out of a total of 11 subjects, 10 were included in the analysis of the far deviants, while only 7 could be included in the analysis of the close deviants. The results for the "no class" group (close deviants) are described in Appendix 3.

Behavioral group	Classification threshold = 0.8	
	FAR dev	CLOSE dev
'two class' AAB – ABB	1,13,18{1}, 11,12,18{2}	1,13,18{1}
'one class' AAN – NBB	5,19, 4, 2{2},7{2}	5,19, 12,7{2}
'no class' ANN – NNB ANB – BBB	2{1},7{1}, 16	2{1},7{1} 2{2},16,4,11,18{2}

TABLE 4.4. Behavioral groups for the analysis of MMN amplitude. Note: {1} and {2} indicate data from subjects that were tested with both sets. Black indicates the data included in the statistical analysis.

4.3.3 Analysis and Results

A two-way ANOVA was performed on the indexes calculated for each subject and each deviant. The two factors were position of the deviant (two levels, inside and outside the class of the standard) and behavioral group (two levels, corresponding to the first two behavioral groups in Table 4.4). Normality of these data were assessed using a Lilliefors test (far deviants: lstat = 0.1382, p-value > 0.2; close deviants: lstat = 0.1396, p-value > 0.2); homogeneity of variances were assessed using a Bartlett test (far deviants: B = 1.914, df = 3, p-value = 0.590; close deviants: B = 2.947, df = 3, p-value = 0.400).

For the far deviants (Figures 4.11 and 4.12) there was no significant interaction and no significant main effect of position or behavior ($df_{BEH} = 1$, $F_{BEH} = 0.19$, $p_{BEH} = 0.67$; $df_{POS} = 1$, $F_{POS} = 0.07$, $p_{POS} = 0.79$; $df_{interaction} = 1$, $F_{interaction} = 0.95$, $p_{interaction} = 0.35$). A two tailed t-test showed that the MMN change indexes of all far deviants combined were around zero (df =19, t = –1.29, p-value = 0.21), indicating that the amplitude of the MMN did not change with training in any condition.

For the close deviants (Figures 4.11, 4.12), there was no significant interaction and no significant main effect of position or behavior (df_{BEH} = 1, F_{BEH} = 1.85, p_{BEH} = 0.20; df_{POS} = 1, F_{POS} = 1.59, p_{POS} = 0.24; $df_{interaction}$ = 1, $F_{interaction}$ = 0.01, $p_{interaction}$ = 0.91). A two tailed t-test showed that the MMN change index of all close deviants combined increased with training (df = 13, t = 2.7580, p = 0.0163). This suggests that the effect of category training was to increase the perceptual salience of close deviants with no regard to their category membership.

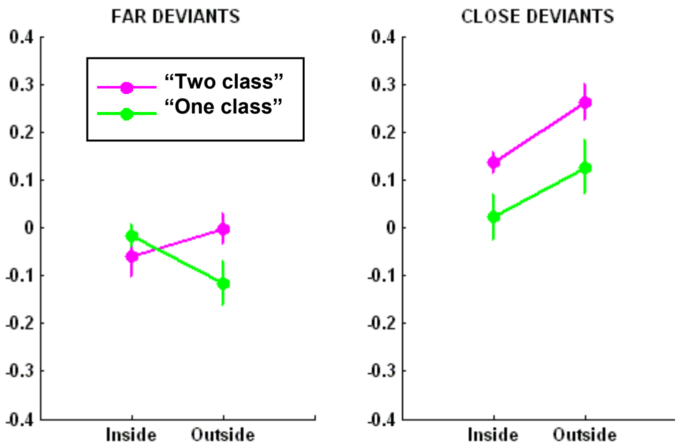


FIGURE 4.11. Plot of the mean normalized indexes for the two deviants. The colors indicate the two behavioral groups. Pink ="two class", Green = "one class".

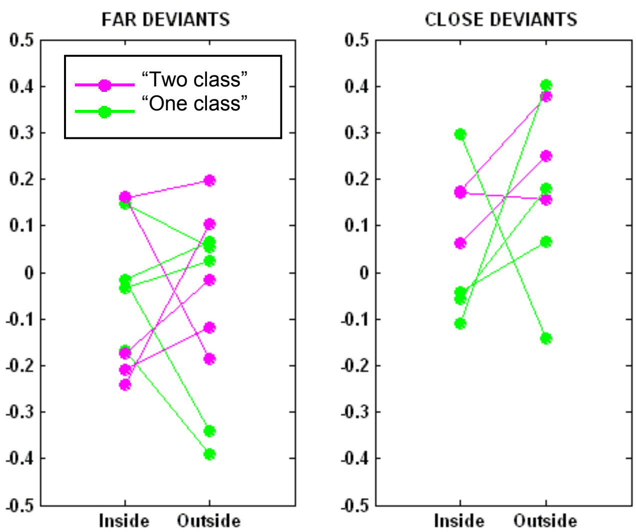


FIGURE 4.12. Plot of the normalized index for the two deviants for each single subject. Colors indicate the two behavioral groups. Pink = "two class", Green = "one class".

4.3.4 Discussion

Far deviants

On average the amplitude of the MMN did not change with training for the far deviants. There was no consistent effect of position or behavioral grouping on the amplitude of the MMN, and the indexes in all conditions were around zero. These results are in agreement with the perceptual auditory staircase data from the same subjects showing that fine perceptual discrimination abilities did not change as a result of training. The sounds used in the perceptual and EEG experiments were not identical but were extremely similar to each other (see Figures 3.7, 3.8 and 4.13).

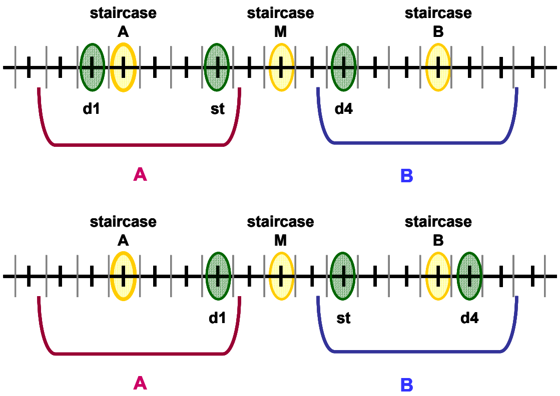


FIGURE 4.13. Position, along the pitch dimension, of the standards and far deviants for the two oddball conditions (green) and of the central frequencies in the set of the auditory staircases (yellow).

This suggests that there was no direct effect of categorization learning on the perceptual abilities of the subjects, in spite of the fact that the subjects learned to perform category membership assignments quite well. In the speech literature some studies (Aaltonen et al., 1987; Sams et al., 1990; Sharma et al., 1993; Maiste et al., 1995) have previously reported an equally strong MMN for items within the same phonemic category and items across categories.

It is also possible that the process of category learning utilized here may require more time and experience to make it sufficiently "automatic" to effectively warp auditory perceptual space. In this case the training provided in the present experiment, although sufficient to yield excellent identification accuracy, may have been insufficient to warp auditory frequency space. However, the training used in the present experiment was much more extensive (on average, 8 sessions) than those used in prior CP training experiments (e.g. Goldstone, 1994; Guenther et al. 1999; Livingstone et al., 1998).

A third possibility is that an influence of categorization was not observed because of some sort of ceiling effect. Since the far deviants are five semitones away from the standard, they might have already been perceived very differently before training. This could have left no room for any increase in MMN amplitude. On the other hand, it could be that categorization does not have an effect on larger scale distances, and only warps perceptual spaces "locally" (for pairs of sounds very close to each other,) either across the borders between categories or within the same category. This could account for the lack of an "acquired similarity" effect at these perceptual distances. The classical categorical perception experiments usually compared the perception to pairs of sounds very close to each other along an identification curve, and thus would not have seen the phenomena that the present experiment has documented for far deviants.

Close deviants

For those subjects that correctly classified the sounds in the oddball task, the "local" perceptual mapping of pitch relevant to the classification seemed to get expanded by training, but the MMN of the close deviants got bigger irrespective of their position relative to the category boundaries. This result is in agreement with Goldstone's (1994) work on visual categorization learning, which found that the stimulus dimension relevant for categorization always got expanded (compared to a control group, which was tested with no training, as Goldstone did not use a pre-test / post-test paradigm like the one used here)[1].

The MMN of the "outside" deviant did not increase after training more than the MMN of the "inside" deviant (see Figures 4.11 and 4.12). Future studies will need to resolve whether such a warping effect would become more evident with longer training and experience (and perhaps practice with more difficult and demanding classification tasks).

A possible explanation for the results of this experiment could be that general learning processes (Amitay, Irvin & Moore, 2006) based on exposure to and training with these sounds, rather than specific categorization learning, influenced the subjects' discrimination abilities. Subjects showed an increase in the perceptual distances between close levels of pitch for the sounds in the regions that were experienced during training, which also generalized to the regions that were not experienced irrespective of where they placed the border between the categories.

4.4 Summary and Final Discussion

This thesis addressed the perceptual effects of auditory category learning. Categorization is the ability to recognize physically different objects as members of the same group. Categorical perception effects arise when objects that are assigned to different categories are perceived as being more different than objects that are assigned to the same category, even when the physical distance between them is the same. Conceptually, when new categories are formed for objects lying on a smooth continuum, the perceptual space is thought to undergo a warping to bring items in the same class closer to each other and/or to bring items from different classes further from each other. Such effects have been hypothesized for phonemic contrasts, which arise or are modified by language learning.

The ways in which category learning might shape perception were investigated here by means of behavioral and psychophysical tests together with electroencephalographic (EEG) recordings of brain activity. Two main questions were asked: whether new and reliable perceptual categories can be established for synthetic sounds varying along two dimensions (only one of which is relevant to categorization) through short-term categorization training; and second, whether such learning results in categorical perception, as assessed by perceptual discriminability and by a measurement derived from EEG recordings that has been empirically associated with perceptual

[1] Another difference between this visual work and the work reported here is that Goldstone tested only the very same items that subjects were trained with; here the expansion involves tests with both trained and untrained items.

discrimination (called mismatch negativity [MMN]). Classification testing showed that training was effective in teaching all subjects how to classify the sounds. Discriminability measures did not support any change in finer perceptual abilities, and a significant increase in the amplitude of the MMN for comparisons involving close deviants (2.5 semitones apart) indicated that, after training, these sounds became perceptually further apart for the subjects, independently of their position with respect to the category boundary. The MMN amplitude for comparisons involving far deviants (5 semitones apart) was unchanged by training, suggesting that the perceptual warping effect of training was specific to perceptual comparisons over small intervals along the trained dimension.

Particular care was taken with the choice of sound characteristics and the selection of subjects. These experiments used synthetic sounds, varying along the ecologically relevant dimensions of pitch and timbre. The final stimuli were selected from two original sets that were characterized according to their timbre perception. First, to assess timbre linearity, a group of subjects was asked to rate differences in stimuli that varied only along the timbre dimension in the two sets. Since neither of the two sets was linear, the stimulus synthesis parameters were modified in order to linearize them according to the results of the rating task. The degree of interference between pitch and timbre was assessed by presented the timbre-linearized sets of sounds to a different group of subjects, who were asked to discriminate stimuli varying simultaneously along both the pitch and timbre dimensions. In both sets of sounds, pitch-timbre interference was present. In one set the interference was less strong, thus it was used as the source of the stimuli for the category learning experiment.

Subjects were selected on the basis of their listening abilities. Subjects were admitted to the categorization experiment if they could judge timbre variations with a minimal interference from pitch variations and vice versa, as assessed by their performance in a discrimination task. About 24 % of all subjects tested (18 out of 75) were able to fulfill these requirements.

A notable feature of this study is its finding of a change in the perceptual scaling of stimuli in the absence of a change in perceptual acuity. Discrimination performance before and after training, as assessed by a set of auditory staircases, showed that the categorization training, although being effective in teaching the subjects to classify the sounds, had no effect on subject's finer perceptual abilities, either in the central regions of the categories or in the region between categories. The results do not support the hypothesis that category learning itself induces a deformation of the perceptual space around categories, because although training increased the amplitude of the MMN brain response to a comparison between tones 2.5 semitones apart, there was no difference in the increase of the 2.5-semitone MMN amplitude for within- or between-category pairs of sounds. Thus, the present study rigorously documents a case where category learning has not changed perceptual acuity, but has locally warped the apparent scaling of the magnitude of differences in the relevant perceptual dimension. Interestingly, this local warping appears generalized and does not finely mirror the specific structure of the categories in a way that would make it conform to classical notions of categorical perception.

It is unlikely that the perceptual warping found here fails to follow the tenets of classical categorical perception because the training regime was ineffective or too short. Although performance was not at chance level in the first session of training, indicating that the categorization task was not extremely difficult, a significant improvement, both in terms of accuracy and reaction time, was achieved by all subjects. After training they all learned to categorize the sounds at high levels of accuracy. Subjects also acquired perceptual expansions along the pitch dimension, as shown by a general improvement in sensitivity to pitch differences (measured by MMN amplitude changes). This enhanced sensitivity did not depend on the subjective place and steepness of the border between the categories, and could be an effect of training in that region. In conclusion, these results suggest that CP is not necessary for good category learning.

Pevtzow & Harnad (1997) proposed that CP is not needed in those categorization tasks where pre-training separation between the two categories are already sufficient to master the task, but it might be necessary to accomplish more difficult tasks where items belonging to two different categories are closer and thus more confusable. This suggestion might apply here, as there was a perceptible gap between the two categories, but this possibility also seems unlikely in the light of the non-categorical acquired distinctiveness that was found for the regions around the gap. This same function could equally well be explained by the operation of an experience-dependent general learning mechanism.

The present results are also consistent with previous critiques of CP studies (Huttenlocher, Hedges & Vevea, 2000; Schouten, Gerrits, van Hessen, 2003; Gerrits & Schouten, 2004) which have suggested that a “two-system” theory better accounts for the results of CP experiments. According to these ideas, a quickly decaying sensory-trace, similar to echoic memory, is involved in immediate comparisons between stimuli based on simple acoustic properties, and this needs to interface with a longer-lasting and more abstract memory system in order to deal with categorical decisions about the stimuli. Whenever the discrimination task requires a longer time, such as in the case of the AX task, the properties of the longer-lasting and more abstract memory representations are well-reflected in behavioral performance. ”Fast” decisions prompted by forced-choice tests, or “early” indicators like the MMN elicited using an oddball paradigm could be more dominated by the properties of the early ”simple acoustic” memory representations. The absence of differential MMN changes for within and between category comparisons as a result of training might be a consequence of the fact that many different “divisions of labor” between these two systems may be compatible with good category learning and performance, and so categorical differences themselves may not necessarily be reflected in early acoustic processing. Thus, classical CP theory may offer too narrow a window for a fuller understanding of the processes through which brains achieve the ability to learn and refine perceptual categories.

Appendices

Appendix 1: Point of convergence in the auditory staircase

The calculations for the convergence point of 70.7 % correct answers are as follows.

According to the 2down-1up rule, whenever two correct answers are given in a row, the distance between the two levels in the pair of sounds decreases, while after each single wrong answer, the distance increases. Thus the set of events that trigger a step down, D, includes only the event of two correct trials in a row (*c,c*), while the set of events that trigger a step up, U, includes the case of single wrong answer and the case in which a wrong answer follows a single correct answer (*w* and *c,w*).

Suppose that in a certain trial the pair of sounds (f_0 ; $f_0 \cdot 2^{n/480}$) is presented in random order. The probabilities of giving a correct or a wrong answer are:

$\mathrm{Prob}_n(c) = p$; $\mathrm{Prob}_n(w) = 1 - p$.

Following the rule, after an event from the U set, the next pair would be (f_0 ; $f_0 \cdot 2^{3/2\,(n/480)}$), while after an event from set D, it would be (f_0 ; $f_0 \cdot 2^{1/2\,(n/480)}$). Thus both the step-down and the step-up, in $\log_2$ scale, are half of (n/480):

$\Delta^+(n) = \Delta^-(n) = \Delta(n) = ½ \cdot (n/480)$.

At the theoretical convergence point (Garcia-Perez, 1998 pointed out that the hypothetical theoretical convergence points of staircase procedures may not be reached under all circumstances), the staircase oscillates around a certain sound pitch level, say $f_0 \cdot 2^{N/480}$. The condition for the oscillation could be expressed by equating the walk up to the walk down; in other words, by equating the two steps up and down weighted by their relative probability of occurrence. Assuming that the oscillation is not too large around the level N, which would depend on the size of N, the two steps can be calculated on the level N. In a $\log_2$ scale this can be expressed as:

$$\mathrm{Prob}(U) \cdot \Delta^+ = \mathrm{Prob}(D) \cdot \Delta^-$$

$$[\mathrm{Prob}(w) + \mathrm{Prob}(c,w)] \cdot \Delta(N) = \mathrm{Prob}(c,c) \cdot \Delta(N)$$

$$\mathrm{Prob}(w) + \mathrm{Prob}(c,w) = \mathrm{Prob}(c,c)$$

$$1 - p + p \cdot (1 - p) = p^2$$

$$p^2 = ½$$

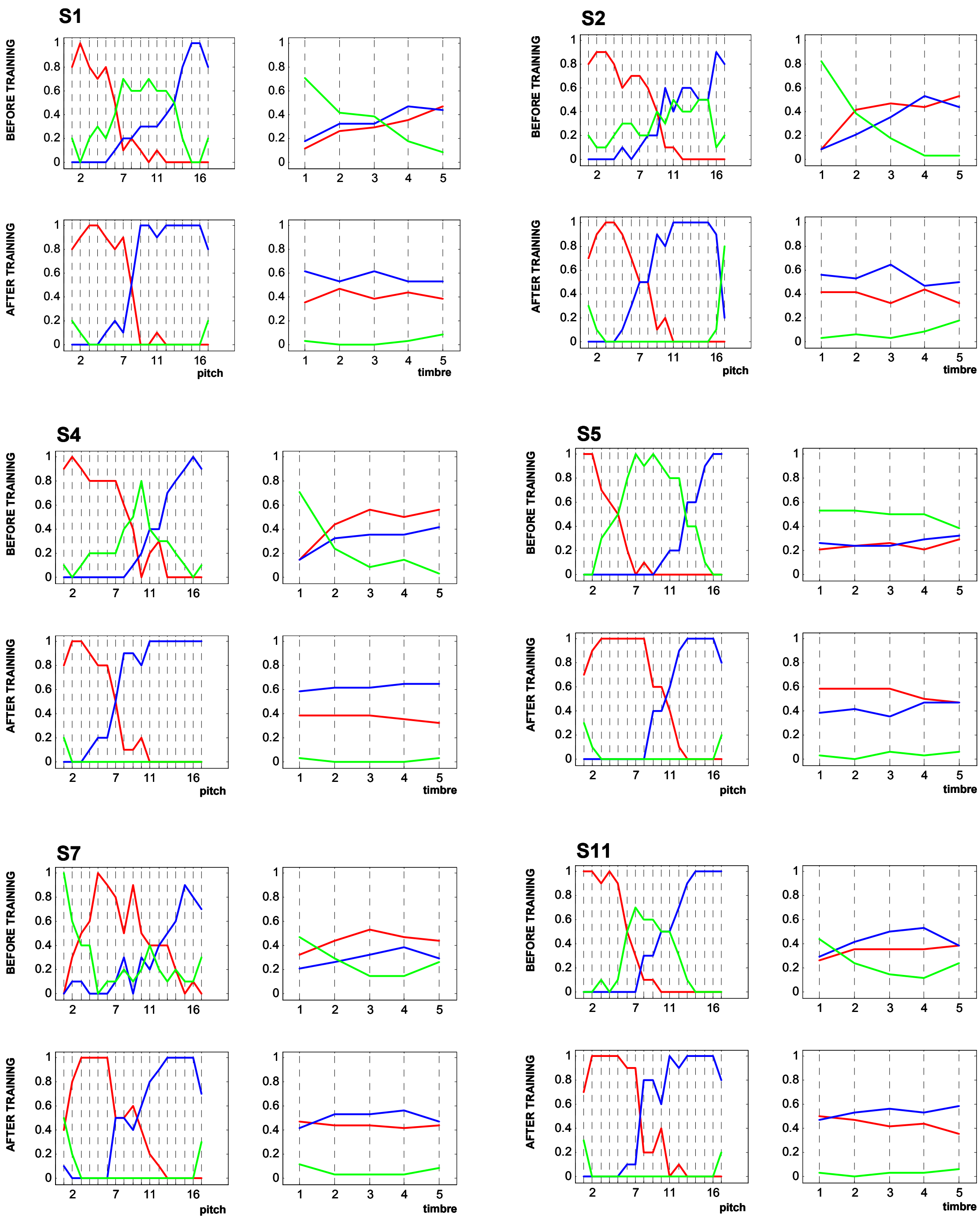
$$p = 0.7071$$

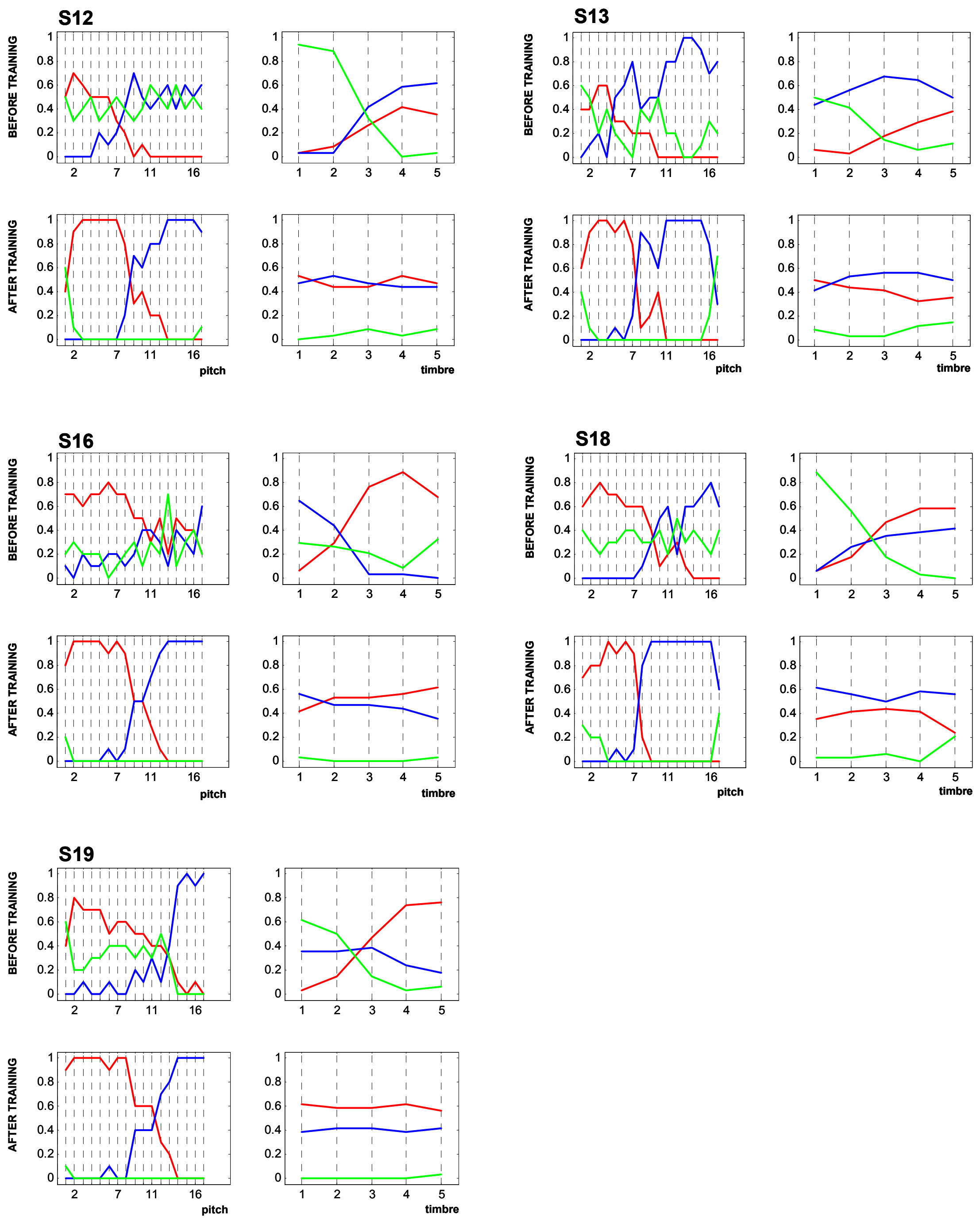
See also Levitt, 1971 and Kaernback, 1991 for similar derivations.

Appendix 2: Single subject plots for the pre-tests and post-tests

A2.1. Identification curves for each subject

The identification curves as a function of the pitch and timbre levels, before and after training are shown here for each subject. The color code is red for the identification curve of category A, blue for category B and green for the answer "no-category".

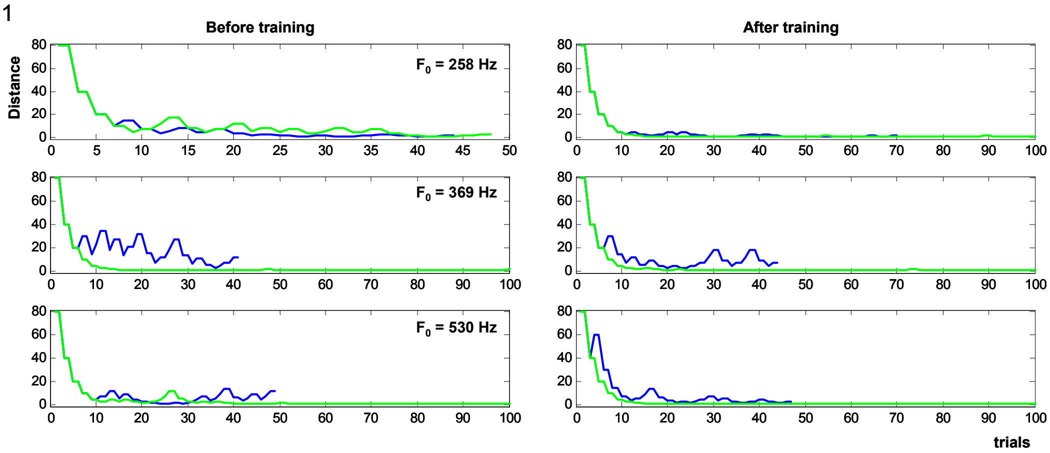




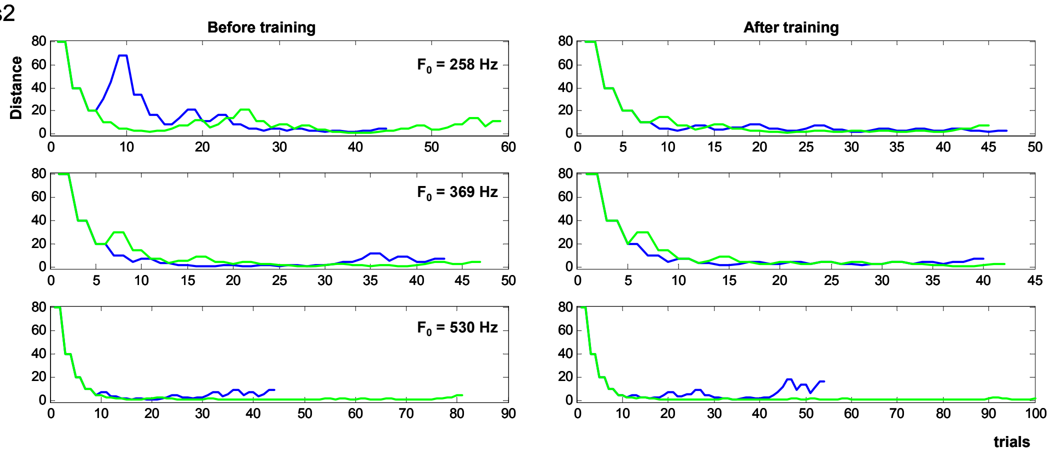
A2.2 Auditory staircases for each subject

The time course of the six staircases is shown for each subject. The staircases where the central sound was always the low frequency sound are shown in green; the ones where the central sound was the high frequency sound are shown in blue.

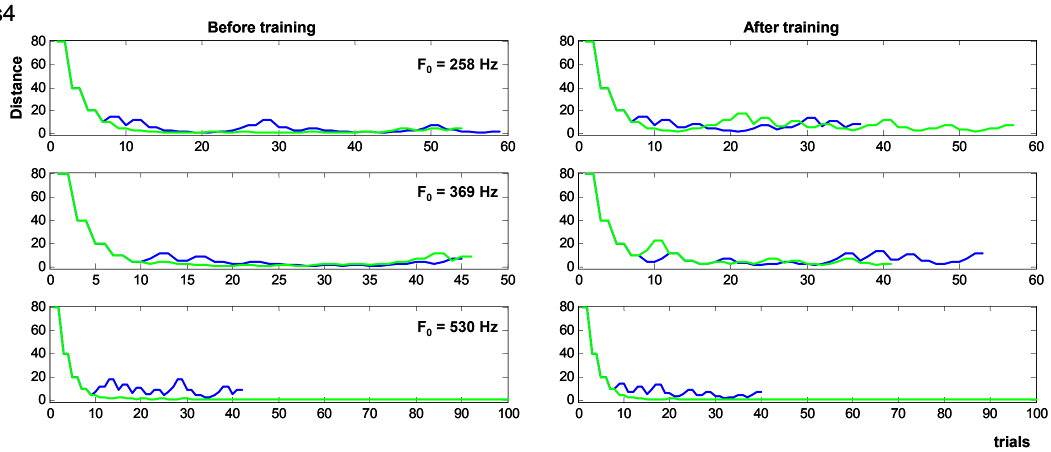
s1


Before training
After training
Distance
F0 = 258 Hz
F0 = 369 Hz
F0 = 530 Hz
trials

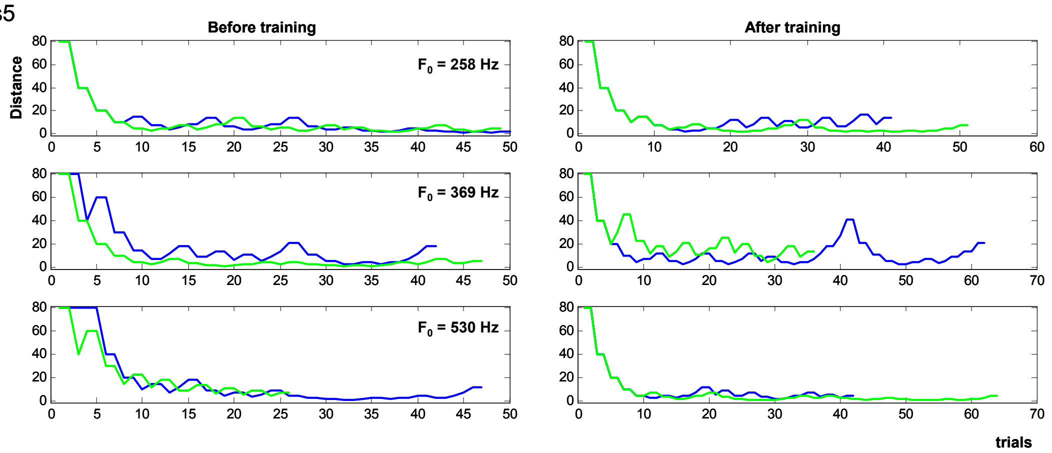
s2


Before training
After training
Distance
F0 = 258 Hz
F0 = 369 Hz
F0 = 530 Hz
trials

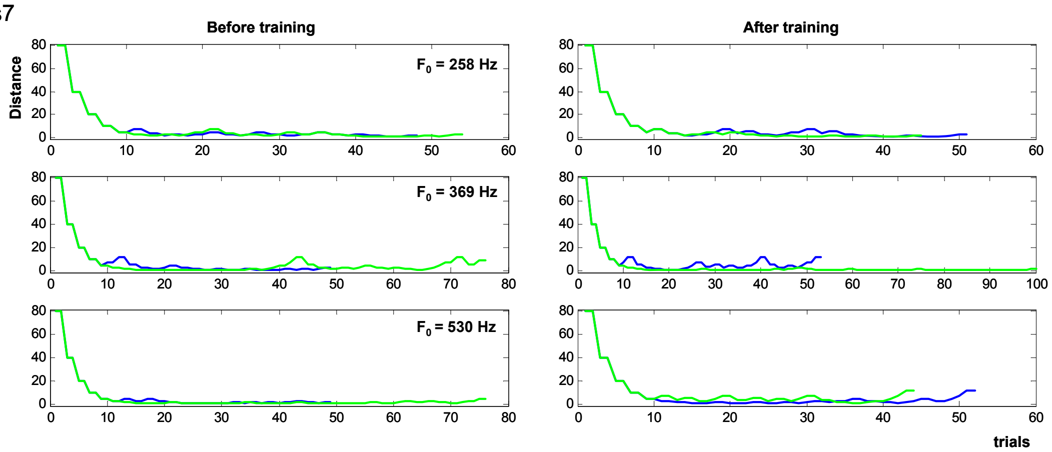
s4


Before training
After training
Distance
F0 = 258 Hz
F0 = 369 Hz
F0 = 530 Hz
trials

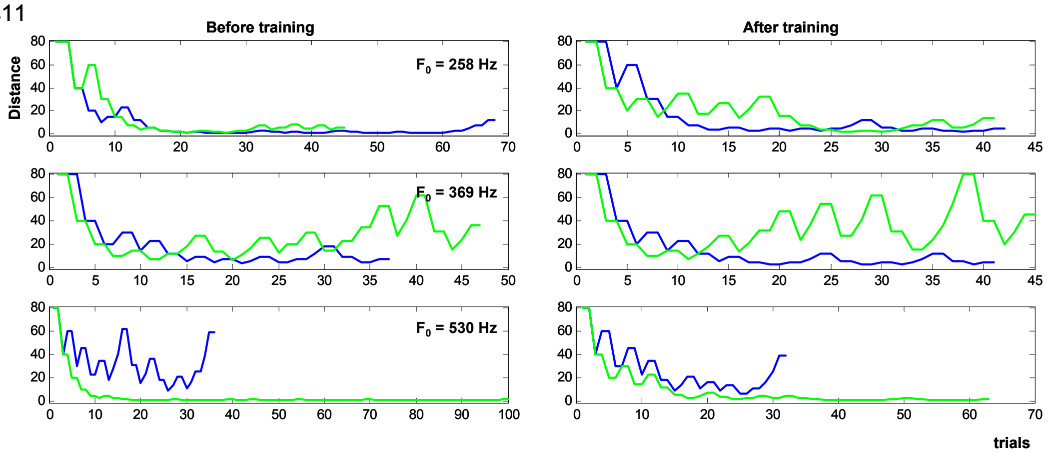
s5


Before training
After training
Distance
F_0 = 258 Hz
F_0 = 369 Hz
F_0 = 530 Hz
trials

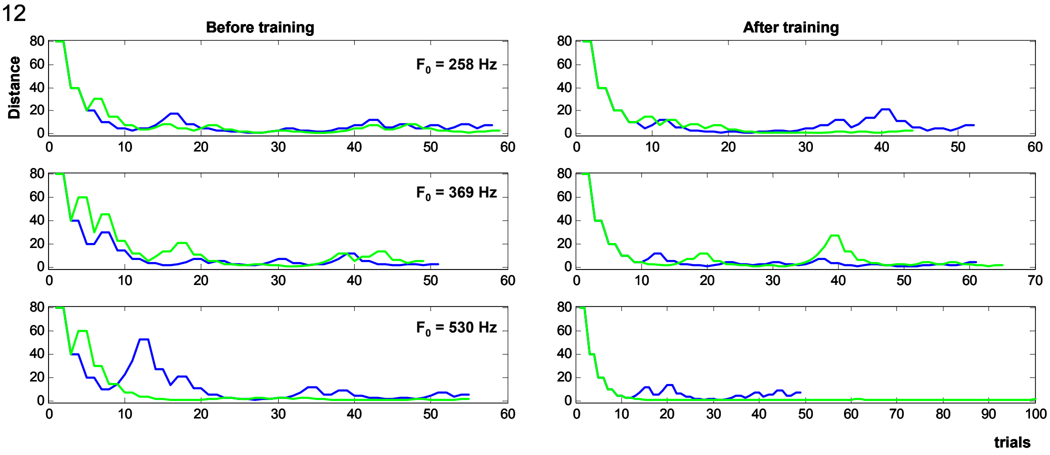
s7


Before training
After training
Distance
F_0 = 258 Hz
F_0 = 369 Hz
F_0 = 530 Hz
trials

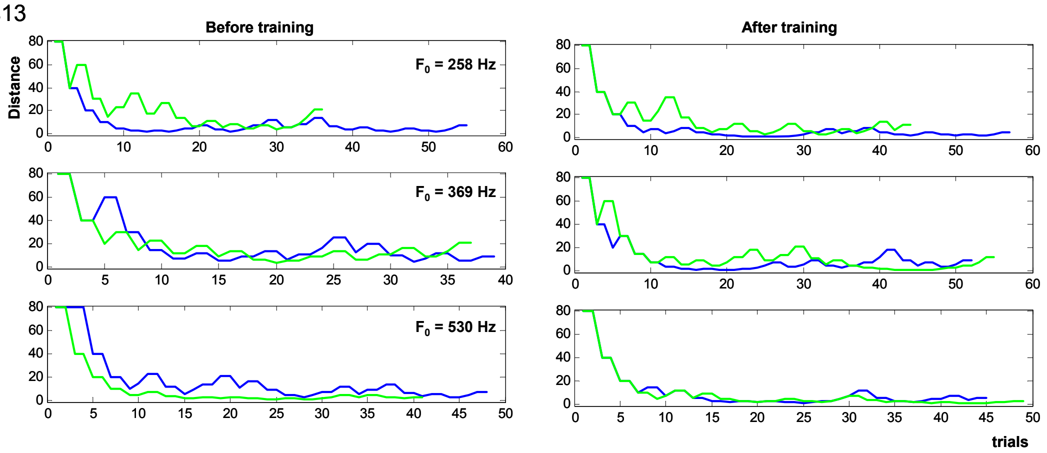
s11


Before training
After training
Distance
F_0 = 258 Hz
F_0 = 369 Hz
F_0 = 530 Hz
trials

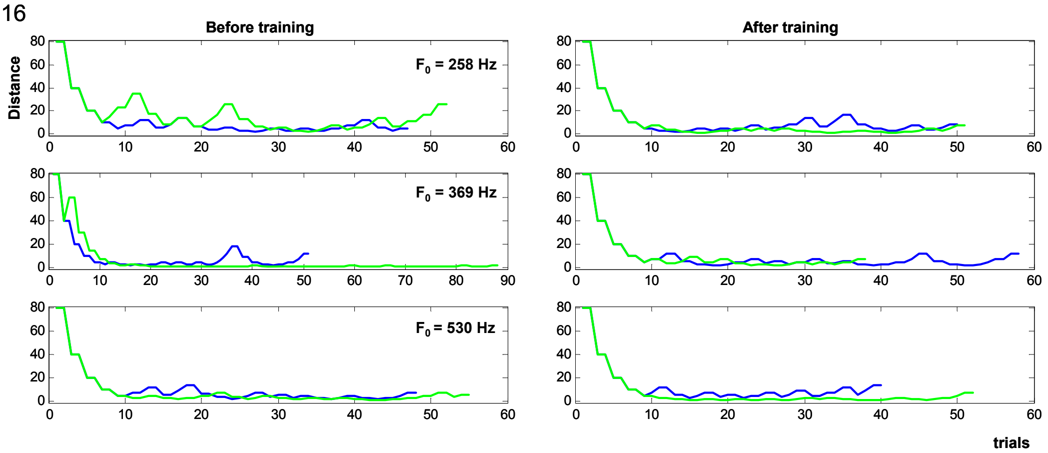
s12


Before training
After training
Distance
F0 = 258 Hz
F0 = 369 Hz
F0 = 530 Hz
trials

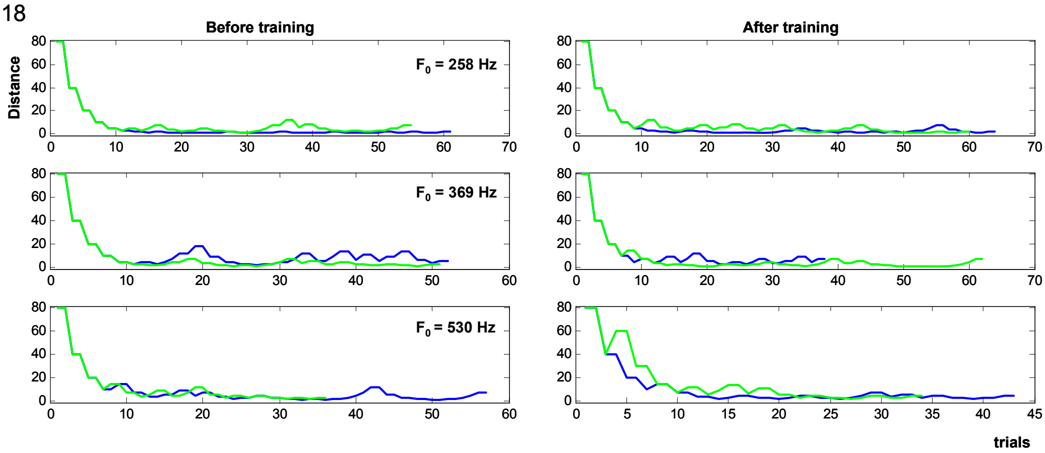
s13


Before training
After training
Distance
F0 = 258 Hz
F0 = 369 Hz
F0 = 530 Hz
trials

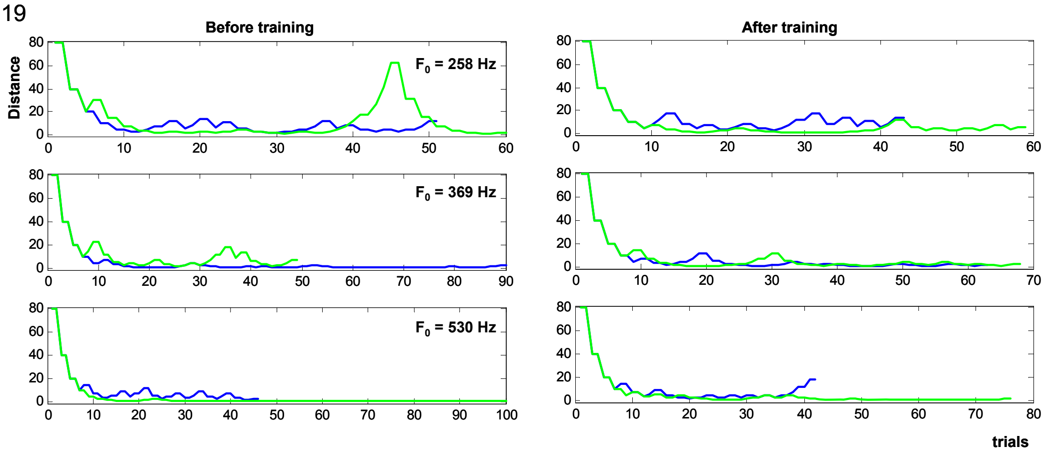
s16


Before training
After training
Distance
F0 = 258 Hz
F0 = 369 Hz
F0 = 530 Hz
trials

s18


Before training
After training
Distance
F_0 = 258 Hz
F_0 = 369 Hz
F_0 = 530 Hz
trials

s19


Before training
After training
Distance
F_0 = 258 Hz
F_0 = 369 Hz
F_0 = 530 Hz
trials

Appendix 3: Results from the 'No-class' Behavioral Group

This Appendix contains an analysis of the MMN amplitude indices for the “No-class” behavioral group for the close deviants.. This group includes subjects 2{1}, 16, 4, 11 (see Table 4.4 and Figure A3.1).

A paired t-test comparing the MMN amplitude indices of the inside and outside deviants showed that there was no significant effect of position (df = 3, t = –2.7107, p = 0.0731). A two tailed t-test showed that the MMN change index of all these close deviants combined did not change with training (df = 7, t = –1.0259, p = 0.3391).

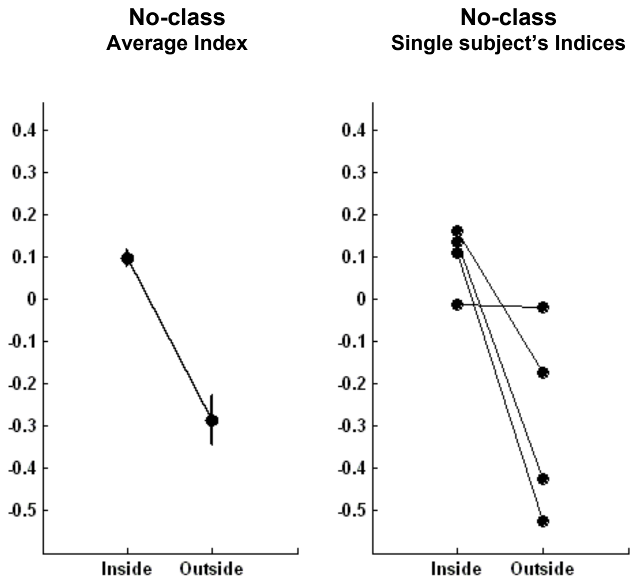


FIGURE A3.1 Left: plot of the mean normalized indices of the “No-class” group for the two close deviants. Right: plot of each single subject’s normalized indices of the “No-class” group for the two close deviants.

In Section 4.3.3 it was shown that for those subjects that correctly classified the sounds in the oddball task, the amplitude of the MMN ***increased*** after training, indicating that the “local” perceptual mapping of pitch was expanded by training. . However, this expansion was seen for all sounds irrespective of their position with respect to the category boundary. In contrast, the ‘no-class’ subjects did not effectively learn the proper placement of the category boundaries with respect to the stimuli used in the EEG experiment, and the amplitude of their MMN did not significantly change with training. Since the “no-class” group is not behaviorally homogeneous, different behavioral trends present in each subject might have simply masked the emergence of any general pattern. On the other hand, these subjects all agree in their failure to have learned the placement of sounds in the region of the ‘close’ oddball stimuli as a result of the same training that produced effective learning in the two other groups of subjects. Thus, their failure to learn about the sounds could be responsible for the lack of an MMN “expansion” of perceptual distances for these sounds. This group could be viewed as a “control” group for the effectiveness of learning: when subjects attend to and learn the attributes of the experimental sounds effectively, they experience a local perceptual expansion for the sounds; when learning is not effective, they do not experience such a local expansion.

Bibliography

Aaltonen O, Eerola O, Hellström A, Uusipaikka E, Lang AH. (1997) Perceptual magnet effect in the light of behavioral and psychophysiological data. *J. Acoust. Soc. Am.,*101(2):1090-105

Aaltonen O, Niemi P, Nyrke T, Tuhkanen M. (1987) Event-related brain potentials and the perception of a phonetic continuum. *Biol. Psychol,.* 24(3):197-207.

Abramson, A.S, & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. *Proceedings of the 6th International Congress of Phonetic Sciences, Prague* (pp. 569-573).

Alho K. (1995) Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear.,* 16(1): 38-51.

American National Standards Institute (1973) *American national psychoacoustical terminology.* S3.20. New York: American Standards Association.

Amitay S, Irwin A, Moore D.R. (2006) Discrimination learning induced by training with identical stimuli. *Nat Neurosci*. 9(11):1446-8.

Beale, J.M. & Keil, F.C. (1995) Categorical effects in the perception of faces. *Cognition*, 57, 217-239.

Berlin, B. & Kay, P. (1969). Basic color terms: Their universality and evolution. Berkeley: University of California Press.

Bornstein MH (1987). Perceptual Categories in Vision and Audition. In S. Harnad (Ed) Categorical perception. Cambridge U. Press.

Bornstein M.H. & Korda N.O. (1984) Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing *Psychol. Res.*, 46:207-222

Burns, E.M., Ward, W.D. (1978). Categorical perception - Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.*, 63, 456-468

Campanella S., Chrysochoos A., Bruyer R. (2001). Categorical perception of facial gender information : behavioural evidence and the face-space metaphor. *Visual Cognition,* 8(2), 237-262.

Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Näätänen R (1998) Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.*, 1(5):351-3.

Cox, T. F. & Cox, M. A. A. (2001) Multidimensional scaling. London: Chapman & Hall.

Csépe, V. (1995) On the origin and development of the mismatch negativity. *Ear Hear.*, *16*, 91–104.

Delorme A., Makeig S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods.,* 134: 9–21.

Diehl RL, Lotto AJ, Holt LL. (2004) Speech perception. *Annu. Rev. Psychol*., 2004;55:149-79

Dooling RJ, Best CT, Brown SD. (1995). Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). *J. Acoust. Soc. Am.,* 97:1839–46

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

Etcoff NL, Magee JJ. (1992) Categorical perception of facial expressions. *Cognition*., 44(3): 227-240.

Fischer, C., Morlet, D., Bouchet, P., Luaute, J., Jourdan, C., Salord, F., (1999) Mismatch negativity and late auditory evoked potentials in comatose patients, *Clin. Neurophysiol*., 110, 1601–1610.

Fischer, C., Morlet, D. & Giard, M. (2000) Mismatch negativity and N100 in comatose patients. *Audiol. Neurootol.,* 5, 192–197

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*, 171-189.

García-Pérez, M. A. (1998) Forced-choice staircases with fixed step sizes: Asymptotic and small sample properties. *Vision Research*, 38,1861-1881.

Garrido K., Hisagi M., and Shafer V. (2004) ERP indices of speech processing in bilinguals. *J. Acoust. Soc. Am.*, Vol. 115, No. 5, Pt. 2.

Gerrits, E., & Schouten, M.E.H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics,* 66: 363-376.

Gibson, E.J. (1963). Perceptual learning. *Annual Review of Psychology*, 14: 29-56.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *J. Exp. Psychol. Gen.,* 123: 178-200.

Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125-157.

Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology*, 49: 585-612.

Grey, J.M. (1977) Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.,* 61, 1270-1277.

Guenther, F.H., Husain, F.T., Cohen, M.A., & Shinn-Cunningham, B.G. (1999). Effects of categorisation and discrimination training on auditory perceptual space. *J. Acoust. Soc. Am.*, 106: 2900-2912.

Handel, S. (1995) Timbre perception and auditory object identification. In B.C.J. Moore (ed.), Hearing (pp. 425-461). New York: Academic Press.

Hazan V. & Barrett S. (2000) The development of phonemic categorization in children aged 6+12 *Journal of Phonetics* 28, 377:396

Heeren, W. & Schouten, M.E.H. (2008). Perceptual development of phoneme contrasts: how sensitivity changes along acoustic dimensions that contrast phoneme categories. *J. Acoust. Soc. Am.*, 124(4), 2291-2302.

Heider, E. R. (1972) Universals in color naming and memory. *Journal of Experimental Psychology,* 95, 10-20.

Hisagi M., (2007) Perception of Japanese temporally-cued phonetic contrasts by Japanese and American English listeners: Behavioral and Electrophysiological Measures. Unpublished doctoral dissertation, City University of NY Graduate School and University Center.

Houtsma A. J. M. **(**1997) Pitch and timbre: Definition, meaning and use. *Journal of New Music Research*, 26, 2, 104 - 115.

Huttenlocher J, Hedges LV, Vevea JL. (2000) Why do categories affect stimulus judgment? *J. Exp. Psychol. Gen.*, 129(2):220-41.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87: B47 - B57.

Jacobsen, T. & Schröger, E. (2001) Is there pre-attentive memory-based comparison of pitch? *Psychophysiology,* 38, 723–727.

Jamieson, D.G., & Morosan, D.E. (1986). Training non-native speech contrasts in adults: acquisition of the English /D-T/ contrast by francophones. *Perception & Psychophysics*, 40: 205-215.

Kaernbach, C. (1991) Simple adaptive testing with the weighted up–down method. *Perception & Psychophysics*, 49, 227-229.

Kraus, N., McGee, T., Carrell, T., King, C., Littman, T., & Nicol, T. (1994). Discrimination of speech-like contrasts in the auditory thalamus and cortex. *J. Acoust. Soc. Am.*, *96*, 2758–2768.

Kraus, N., McGee, T., Carrell, T., King, C., Tremblay, K. and Nicol, T., (1995) Central auditory system plasticity associated with speech discrimination training. *J. Cogn. Neurosci.,* 7, pp. 25–32.

Krumhansl C. L. & Iverson P. (1992) Perceptual Interactions Between Musical Pitch and Timbre. *Journal of Experimental Psychology: Human Perception and Performance,* 18(3): 739-751.

Kuhl P.K. (1981). Discrimination of speech by nonhuman animals: basic auditory sensitivities conducive to the perception of speech-sound categories. *J. Acoust. Soc. Am.,* 95:340–49

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. deBoysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), Developmental neurocognition: Speech and face processing in the first year of life (pp. 259-274). Dordrecht, Netherlands: Kluwer Academic Publishers.

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. *J. Acoust. Soc. Am*, 63, 905-917.

Kuhl, P.K., & Padden, D.M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics,* 32: 542-550.

Kuhl, P. K. & Rivera-Gaxiola, M. (2008). Neural substrates of early language acquisition. *Annual Review of Neuroscience,* 31, 511-534

Kuhl, P K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science,* 9, F13-F21

Kujala T., Tervaniemi M., Schröger E. (2007) The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol. Psychol*., 74(1):1-19.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review,* 72: 275–309

Lawrence, D.H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology,* 39: 770-784.

Levin, D.T., & Angelone, B.L. (2002). Categorical perception of race. *Perception,* 31 , 567-578 .

Levitt, H. (1971). Transformed up–down methods in psychophysics. *Journal of the Acoustical Society of America*, 49, 467-477.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review,* 74: 431–461.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 4, 175-195.

Liberman, A.M., Harris, K.S., Hoffman, H.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. (1970). The voicing dimension: some experiments in comparative phonetics. *Proceedings of VIth ICPhS, Prague*, 563-567.

Lively, S. E., Logan, J. S. & Pisoni, D. B. (1993) Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories.*J. Acoust. Soc. Am.,* 94, 1242-1255.

Livingston, K., Andrews, J. and Harnad, S. (1998) Categorical Perception Effects Induced by Category Learning*. Journal of Experimental Psychology: Learning, Memory and Cognition*, 24 (3). pp. 732-753.

MacKain, K.S., Best, C.T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics,* 2: 369-390.

Maiste, A., Wiens, A., Hunt, M., and Picton, T. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear Hear.,* 16, 68–90.

McAdams, S., & Cunibile, JC. (1992) Perception of timbral analogies. *Philosophical Transactions of the Royal Society*, vol. 336, London Series B 1992.

Melara R. D. & Marks L. E. (1990a) HARD and SOFT interacting dimensions: Differential effects of dual context on classification. *Perception & Psychophysics,* 47: 307-325.

Melara R. D. & Marks L. E. (1990b) Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics,* 48: 169-178.

Melara R. D. & Marks L. E. (1990c) Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance,* 16: 398-414.

Mervis C. B. & Catlin J., Rosch E. (1975) Development of the Structure of Color Categories *Developtnemal Psychology,* Vol. I I, No. I, 54-60

Miller JD, Wier CC, Pastore RE, Kelly WJ, Dooling RJ. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception. *J. Acoust. Soc. Am.,* 60:410–17

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J.J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 18: 331-340.

Moore B.C.J., Glasberg B.R. and Peters R.W. (1985) Relative dominance of individual partials in determining the pitch of complex tones, *J. Acoust. Soc. Am.*, 77: 1853-1860.

Näätänen, R. (1992) Attention and Brain Function. Lawrence Erlbaum, Hillsdale, New Jersey.

Näätänen, R. (2001) The perception of speech sounds by the human brain as reflected by the mismatch negativity MMN and its magnetic equivalent MMNm *Psychophysiology, 38* 1–21

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature,* 30; 385 (6615): 432-4.

Näätänen R., Picton, T. (1987) The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychopchysiology*, 24, 4, 375:425.

Näätänen, R., Schroger, E., Karakas, S., Tervaniemi, M., and Paavilainen, P. (1993) Development of neural representations for complex sound patterns in the human brain. *NeuroReport* 4, 503–506.

Näätänen, R., & Winkler, I. (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *6*, 826–859

Nosofsky RM. (1986) Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.,* 115(1): 39-61.

Özgen, E; Davies, IRL. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *J. Exp. Psychol. Gen.,* 131 (4): 477-493.

Pevtzow, R. & Harnad, S. (1997) Warping Similarity Space in Category Learning by Human Subjects: The Role of Task Difficulty. In: Ramscar, M., Hahn, U., Cambouropolos, E. & Pain, H. (Eds.) *Proceedings of SimCat 1997: Interdisciplinary Workshop on Similarity and Categorization.* Department of Artificial Intelligence, Edinburgh University: 189 - 195.

Pilling, M; Wiggett, A; Ozgen, E; Davies, IRL. (2003). Is color "categorical perception" really perceptual?. *Memory & Cognition,* 31 (4): 538-551

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13: 253-260.

Pisoni, D.B. (1991). Modes of processing speech and nonspeech signals. In I.G. Mattingly & M. Studdert-Kennedy (eds.): Modularity and the motor theory of speech perception, 225-238. Hillsdale, NJ: Lawrence Erlbaum Associates.

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. (2009) Neural correlates of categorical perception in learned vocal communication.*Nat. Neurosci.,* 12(2):221-8.

Reber PJ, Stark CE, Squire LR. (1998) Cortical areas supporting category learning identified using functional MRI.*Proc Natl Acad Sci U S A.,* 20;95(2):747-50.

Roberson, D., Davies, I. & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.,* 129: 369–398..

Rosch, E. (1973) Natural categories. *Cognitive Psychology*, *4,* 328-350.

Sallinen, M., Kaartinen, J., & Lyytinen, H. (1996). Processing of auditory stimuli during tonic and phasic periods of REM sleep as revealed by event-related brain potentials. *Journal of Sleep Research*, *5*, 220–228.

Sams, M., Aulanko, R., Aaltonen, O., and Näätänen, R. (1990) Eventrelated potentials to infrequent changes in synthesized phonetic stimuli. *J. Cognitive Neuroscience,* 2, 344–357.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41: 71-80.

Shafer VL, Schwartz RG, Kurtzberg D. (2004) Language-specific memory traces of consonants in the brain. *Cogn. Brain Res*., 18(3):242-54.

Sharma, A., & Dorman, M. F. (1998) Exploration of the perceptual magnet effect using the mismatch negativity auditory evoked potential. *J. Acoust. Soc. Am.,* 104(1):511-7.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.,* 106, 1078–1083.

Sharma A, Kraus N, McGee T, Carrell T, Nicol T. (1993) Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalogr Clin Neurophysiol.* 88(1): 64-71.

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell *sharp*from *flat*. *Perception & Psychophysics,* 21(5), 399-407.

Sinex, D., and McDonald, L. (1989) Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. *J. Acoust. Soc. Am.,* 85, 1995–2004.

Sinex, D., McDonald, L., and Mott, J. B. (1991) Neural correlates of nonmonotonic temporal acuity for voice-onset time. *J. Acoust. Soc. Am.,* 90, 2441–2449.

Sinex, D., and Narayan, S. (1994) Auditory-nerve fiber representation of temporal cues to voicing in word-medial stop consonants. *J. Acoust. Soc. Am.,* 95, 897–903.

Smith D. R. R. and Patterson R. D. (2005) The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J. Acoust. Soc. Am.,* 118.

Soloudre, G. A. (2004) Evaluation of Objective Loudness Meters. Proc. 116th AES Convention.

Soloudre, G. A. & Norcross, S. G. (2003) Objective Measres of Loudness. Proc. 115th AES Convention.

Stevens, K. N., Liberman, A.M., Studdert-Kennedy, M. G., & Ohman, S. E. G. (1969). Cross-language study of vowel perception. *Language and Speech, 12*, 1-23.

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., & Landauer, T.K. (1976). Effects of learning English as a second language on the acquisition of a new phoneme contrast. . *J. Acoust. Soc. Am.,* 59: 448-451.

Sussman, J.E. (1993). Auditory processing in children's speech perception: results of selective adaptation and discrimination tasks. *Journal of Speech and Hearing Research*, 36: 380-395.

Szymanski MD, Yund EW, Woods DL (1999) Phonemes, intensity and attention: differential effects on the mismatch negativity (MMN). *J. Acoust. Soc. Am.,* Dec;106(6):3492-505.

Terhardt, E. (1974) Pitch of pure tones: its relation to intensity *Facts and Models in Hearings,* edited by E. Zwicker and E. Terhardt (Springer, Berlin).

Tiitinen, H., May, P., Reinikainen, K. & Näätänen, R. (1994) Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature,* 372, 90–92

Tremblay, K., Kraus, N., Carrell, T.D. and McGee, T., (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.,* 102, pp. 3762–3773.

Ulanovsky N., Las L., Nelken I., (2003) Processing of low-probability sounds by cortical neurons *Nature Neuroscience,* 6, 391 - 398

Valberg A. (2001) Unique hues: an old problem for a new generation. *Vision Research 41* 1645–1657

Werker, J.F., & Polka, L. (1993). Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics*, 21: 83-101.

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

Wier C.C., Jesteadt W. and Green D.M. (1977) Frequency discrimination as a function of frequency and sensation level, . *J. Acoust. Soc. Am.,* 61 pp. 178–184.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics*, 26: 95-104.

Winkler I, Kujala T, Tiitinen H, Sivonen P, Alku P, Lehtokoski A, Czigler I, Csépe V, Ilmoniemi RJ, Näätänen R. (1999) Brain responses reveal the learning of foreign language phonemes. *Psychophysiology,* 36(5):638-42.

Woldorff, M.G., Hillyard, S.A., Gallen, C.C., Hampson, S.R., Bloom, F.E. (1998) Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex, *Psychophysiology,* 35, 283–292.

Zahid, M. (1996). Examination of a perceptual non-native speech contrast: pharyngealized/non-pharyngealized discrimination by French-speaking adults. *Proceedings of ICSLP, Philadelphia*, 2466-2469.

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics, 26*(5), 384-395.