Cognition 112 (2009) 1–20

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Processing interactions between phonology and melody: Vowels sing but consonants speak

Régine Kolinsky [a,b,*], Pascale Lidji [b,c], Isabelle Peretz [c], Mireille Besson [d], José Morais [b]

[a] Fonds de la Recherche Scientifique – FNRS, 5 rue d'Egmont, B-1000 Brussels, Belgium
[b] UNESCOG, Université Libre de Bruxelles, ULB, Av. F. Roosevelt, 50 CP 191, B-1050 Brussels, Belgium
[c] BRAMS Laboratory, Université de Montréal, 1430 Bd Mont-Royal Ouest, Montréal H2V4P3, Québec, Canada
[d] Institut de Neurosciences cognitives de la Méditerranée, CNRS, 31 chemin J. Aiguier, 13402 Marseille Cedex, France

ARTICLE INFO

Article history:
Received 1 October 2008
Revised 16 January 2009
Accepted 19 February 2009

Keywords:
Music-language processing interactions
Consonant-vowel processing differences

ABSTRACT

The aim of this study was to determine if two dimensions of song, the phonological part of lyrics and the melodic part of tunes, are processed in an independent or integrated way. In a series of five experiments, musically untrained participants classified bi-syllabic nonwords sung on two-tone melodic intervals. Their response had to be based on pitch contour, on nonword identity, or on the combination of pitch and nonword. When participants had to ignore irrelevant variations of the non-attended dimension, patterns of interference and facilitation allowed us to specify the processing interactions between dimensions. Results showed that consonants are processed more independently from melodic information than vowels are (Experiments 1–4). This difference between consonants and vowels was neither related to the sonority of the phoneme (Experiment 3), nor to the acoustical correlates between vowel quality and pitch height (Experiment 5). The implication of these results for our understanding of the functional relationships between musical and linguistic systems is discussed in light of the different evolutionary origins and linguistic functions of consonants and vowels.

1. Introduction

A fundamental issue in human cognition is to determine how the different dimensions of a stimulus combine and interact in processing complex materials. Speech and music are typical examples of such materials, and have been studied not only for their own sake, but also for comparing the cognitive processes involved in each of them. While some authors view music processing as a by-product of language processing (e.g., Pinker, 1997), others argue that music involves specific computational processes (e.g., Peretz, 2006). Songs provide an ideal material to study the relations between language and music, since they naturally combine a musical dimension, the tune, and a linguistic dimension, the lyrics (e.g., Patel & Peretz, 1997). The aim of the present work was to examine whether lyrics and tunes in sung materials are processed independently or in an integrated way. More specifically, we examined the on-line processing independence or integration of the phonological and melodic dimensions of sung materials.

Up to now, studies on songs have mainly investigated the relations between semantics and melody. Depending on the experimental approach and on the materials used, results show either independence (Besson, Faïta, Peretz, Bonnel, & Requin, 1998; Bonnel, Faïta, Peretz, & Besson, 2001) or interactions (Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005; Schön, Gordon, & Besson, 2005). The effect of harmonic congruity on phoneme monitoring seems to suggest that interactive processing of phonology

* Corresponding author. Address: UNESCOG, Université Libre de Bruxelles, ULB, Av. F. Roosevelt, 50 CP 191, B-1050 Brussels, Belgium. Tel.: +32 2 650 42 30; fax: +32 2 650 22 09.
E-mail address: rkolins@ulb.ac.be (R. Kolinsky).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.02.014

and music occurs. In studies of *harmonic priming*,[1] Bigand, Tillmann, Poulin, D'Adamo, and Madurell (2001) manipulated the structural relationship between the last sung chord and the preceding musical context, an eight-chord sung sequence. Results showed faster phoneme monitoring of the last sung vowel when it was sung on the tonic than on the subdominant chord. However, if linguistic and musical domains shared some attentional capacities, music may modulate linguistic processing by modifying the allocation of attentional resources necessary for linguistic computation. Under this view, the effect of harmonic context on phoneme processing arises from general attentional processes rather than from specific music-language dependencies (Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001; see also Poulin-Charronnat et al., 2005). This possibility is supported by recent evidence of similar facilitation from harmonic relatedness with nonlinguistic stimuli, such as geometric shapes (Escoffier & Tillmann, 2008). In addition, Bigand, Tillmann, Poulin, D'Adamo, and Madurell (2001) only used one phoneme category for discrimination, namely vowels (i.e., the /di/-/du/ distinction). Their finding may not generalize to harmonic and phonemic processing as a rule, since vowels and consonants differ in both their acoustical properties and linguistic function.

At the acoustical level, most consonants are characterized by transient acoustic cues typical of formant transitions, whereas vowels are characterized by the relationship between more steady-state frequency information (Delattre, Liberman, & Cooper, 1955; Fry, Abramson, Eimas, & Liberman, 1962; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). These acoustical differences have been associated with different cerebral hemispheres: the processing of rapidly changing acoustic information (e.g., consonants) is more left-lateralized than the processing of stable spectral information (e.g., vowels or music; for reviews, see Poeppel, 2003; Zatorre, Belin, & Penhune, 2002; Zatorre & Gandour, 2008). Therefore, vowels might be more suitable to carry melodic and prosodic information than consonants. This idea is in line with studies on opera singing, which suggest that vowels are more intimately linked to melodic variations of tunes than consonants, as the latter are located at the transition between notes and are sometimes reported as breaking the melodic line (e.g., Scotto di Carlo, 1993). Thus, trained singers tend to shorten consonants and to reduce their articulation (McCrean & Morris, 2005), while vowels are typically lengthened in singing compared to speech (Scotto di Carlo, 2007a, 2007b; Sundberg, 1982).

At the functional level, vowels and consonants may also serve distinct roles in speech (Bonatti, Peña, Nespor, & Mehler, 2007). Statistical learning studies have shown that humans are better at capturing non-adjacent regularities based on consonants than on vowels (Bonatti, Peña, Nespor, & Mehler, 2005; Mehler, Peña, Nespor, & Bonatti, 2006). This suggests that consonants carry lexical information. The specific lexical function of consonants seems to emerge relatively early in human life, given that 20 month-old infants can learn two words that differ by only one consonant, but fail when the distinctive phoneme is a vowel (Nazzi, 2005; Nazzi & New, 2007). In contrast with the lexical function of consonants, vowels are used to extract structural generalizations in artificial languages (Toro, Nespor, Mehler, & Bonatti, 2008). Vowels are thus involved in syntactic computations. They also contribute to grammar and to prosody (Nespor, Peña, & Mehler, 2003; Toro et al., 2008), including indexical prosody that allows for speaker identification (Owren & Cardillo, 2006).

In addition, neuropsychological dissociations have been reported between the ability to produce vowels and consonants in aphasic patients (Caramazza, Chialant, Capasso, & Miceli, 2000), with consonants being more vulnerable than vowels to such impairments (Béland, Caplan, & Nespoulous, 1990; Canter, Trost, & Burns, 1985; for a review, see Monaghan & Shillcock, 2003; but see Semenza et al., 2007). These two classes of speech segments would thus pertain to distinct processing systems.

Comparative human–animal studies further suggest that consonants are more specific to human speech than vowels are. Contrary to humans, New World monkeys (cotton-top tamarins) are only able to extract statistical regularities based on vowels (Newport, Hauser, Spaepen, & Aslin, 2004). Moreover, while monkeys have a steady-state formant perception comparable to the one of humans (Sommers, Moody, Prosen, & Stebbins, 1992), and can learn to discriminate the manner of articulation of consonants (Sinnott & Williamson, 1999), they exhibit problems in learning the place of articulation contrasts. Consequently, they process formant transitions differently from humans (Sinnott & Gilmore, 2004). On the production side, nonhuman primates can produce harmonic sounds very similar to vowels in order to provide indexical information about sex, age, identity, emotion, etc. (Owren, Seyfarth, & Cheney, 1997; Rendall, Rodman, & Emond, 1996). However, only humans have elaborated the supralaryngeal articulations that, by inserting consonants into the vocalic carrier (MacNeilage & Davis, 2000), allow the emergence of a rich set of meaningful contrasts.

In summary, learning and developmental research support the notion that vowels and consonants subtend different linguistic functions, with consonants being more tied to word identification, while vowels essentially contribute to grammar and to prosody. In addition, neuropsychological dissociations show that the processing of these two classes of speech segments is dissociable by brain damage. Furthermore, comparative human–animal studies suggest that vowels may be less specific to speech than consonants. As a consequence, vowels may be more intricately linked than consonants to other non-linguistic auditory dimensions, like melody.

To our knowledge, this hypothesis has only been tested in speech using auditory adaptations of the *speeded classification* tasks designed by Garner (e.g., Garner, 1974, 1978a, 1978b; see also Lidji (2007), for a recent review in the song domain). In these tasks, participants are asked to classify spoken syllables according to their values on a previously

[1] Harmonic priming reflects the musical function of the chord in the previous key context. Chords that do not belong to the musical key context (e.g., Bharucha & Stoeckig, 1986) or that are less stable in the key context (such as a *subdominant* chord compared to a *tonic* chord) are less primed by the context, resulting in slower processing even in musically naïve participants (e.g., Bigand & Pineau, 1997).

specified target dimension. This dimension could be, for example, the pitch level (manipulated through the vowel fundamental frequency, F0) or the identity of the initial consonant (e.g., /bæ/ or /gæ/; Wood, 1974, 1975). Three conditions constitute the *filtering* and *redundancy tests* (Ashby & Maddox, 1994) that aim to check whether irrelevant orthogonal variations on one dimension (e.g., identity of the consonant) influence the processing of the other, namely the target dimension (e.g., pitch). Variations on the two dimensions can be either redundant (i.e., correlated, e.g., when all /bæ/ syllables have a low pitch and all /gæ/ syllables have a high pitch) or orthogonal (when both /bæ/ and /gæ/ syllables can be either low or high). Comparing sorting times and performance with a *baseline* control test (also called a *standard* or *discrimination* task, e.g., Garner, 1981), where only one dimension is varied (e.g., just the consonant, with only high /bæ/ and /gæ/ syllables, or just pitch, with only high and low /bæ/), allows one to evaluate the participants' attentional filtering capacities. Indeed, if processing of the target dimension entailed processing of the non-target dimension, participants would be unable to filter out irrelevant variations. Hence, their performance would be poorer (e.g., slower Reaction Times, RTs) in the filtering test[2] than in the baseline tests, an effect referred to as *Garner interference* (e.g., Pomerantz, 1983).[3]

With this speeded classification paradigm, the interactions between segmental (phonemes) and suprasegmental (pitch or pitch contour) dimensions in speech have been shown to be modulated by the nature of the phonemes. While pitch classification was not affected by the consonantal variations described above (Wood, 1974, 1975), consonant classification was slowed down by variations in pitch. In contrast, when the segmental task concerned vowel quality (e.g., /bɑ/ vs. /bæ/) rather than consonants, mutual and symmetric interference between the segmental dimension, and either the pitch or the loudness dimension, was reported (Miller, 1978). These data seem to support the idea that vowels and consonants have different relationships with pitch. According to Melara and Marks (1990), vowel and pitch are processed by the same general auditory mechanisms, while consonants are processed at a later level, a phonetic one.

However, these results cannot be generalized to lyrics and tunes in songs. While static pitch levels (i.e., synthetic syllables recorded at a constant F0: 104 or 140 Hz) were used in these studies (Miller, 1978; Wood, 1974, 1975), music (including songs) as well as speech intonation and lexical tones are characterized by pitch changes. Using dynamic tonal contours in speeded classification, Repp and Lin (1990) observed mutual interference between segmental (consonant or vowel) and tonal information in Mandarin Chinese. More crucially, in English listeners, Lee and Nusbaum (1993) observed mutual interference between consonantal and pitch information for dynamic tonal contours but asymmetrical interference for static pitch levels. Thus, contrary to what was observed with static pitch levels, both vowels and consonants interact with speech tonal contours.

Yet, there are several shortcomings in these studies. In most of them, processing interactions between dimensions were assessed only by examining the interference pattern, which may merely reflect the listeners' inability to pay attention selectively to the target dimension (e.g., Thibaut & Gelaes, 2002). According to Garner (1974), a demonstration of *integrality* of multidimensional stimuli, namely of integrated, *holistic* processing, requires not only the occurrence of interference, but also that correlated variations on the non-target dimension lead to a benefit or *redundancy gain*. Indeed, in the redundant condition, when the dimensions are processed in a *unitary fashion* (Grau & Kemler-Nelson, 1988), the perceptual distance between the whole stimuli is enhanced according to a Euclidean metric of (dis)similarity. By contrast, for separable dimensions, (dis)similarity is based on a *city-block* metric in which (dis)similarity between multidimensional stimuli is additive (Torgerson, 1958), and hence no gain is expected in the redundant condition. To our knowledge, the only study that also used correlated stimuli (i.e., the redundancy test) to examine linguistic tones and segmental information interactions was that of Repp and Lin (1990). Unfortunately, in this study, there was a difference in the relative discriminability of the two dimensions (all participants discriminated tones more poorly than segments), which is known to modulate the patterns of dimensional interaction (Garner, 1974; Garner & Felfoldy, 1970). In addition, even results obtained with speech tonal contours are unlikely to generalize to interactions between other auditory dimensions such as lyrics and tunes of songs. For instance, it has been shown that "consonants" and "vowels" of synthesized consonant–vowel (CV) stimuli are processed as integral dimensions when listeners consider them as linguistic, but separately when considered as a mix of noise and tone (Tomiak, Mullennix, & Sawusch, 1987).

In the present work, we adapted the speeded classification paradigm to study the processing of the phonological and melodic dimensions of consonant–vowel bisyllabic (CVCV) nonwords sung on two-tone melodic intervals. Musically untrained participants were presented with speeded classification tasks, using either natural (Experiments 1, 2, 3 and 4) or synthesized (Experiment 5) sung syllables. The speeded classification tasks of Experiment 1, 2, 3 and 5 included the filtering, redundant and baseline conditions that constitute the filtering and redundancy tests. In all these conditions, participants were required to respond according to the identity of either the "lyrics" (the nonword) or the "tune" (the melodic interval). We contrasted materials in which the nonwords differed by their middle consonant, either a stop (Experiment 2) or a nasal (Experiment 3), to materials in which they differed by their final vowel (Experiments 1, 3 and 5). The rationale for contrasting these materials was to test the hypothesis that vowels and consonants involve different processes that may have different relationships with melodic processing.

[2] This nomenclature follows the one proposed by Posner (1964) see also Ashby and Maddox (1994), but the same situation is also often called *orthogonal classification* (e.g., Patching & Quinlan, 2002).

[3] Garner interference should be distinguished from Stroop-like congruency effects (Stroop, 1935): rather than arising from the content of an irrelevant dimension that is present on every trial and that is assigned an incompatible response, it arises from variations on the irrelevant (but not necessarily incongruent) dimension across trials.

In Experiment 1, we examined vowels and intervals processing. If vowels and intervals constituted interactive dimensions, an integrality pattern – interference cost and redundancy gain – was expected. In Experiment 2, the nonwords differed by their middle, voiceless stop consonant. If consonants were more speech-specific and provided poor melodic support, a separability pattern was predicted.

The aim of Experiment 3 was to generalize results beyond the case of stop consonants, as well as to a new vowel contrast. By definition, consonants and vowels differ in their acoustics. They also differ in the physics of their production, because only consonants are produced by either a partial or total constriction of the upper vocal tract. However, some consonants are more sonorous – and hence more vowel like – than others. Indeed, in all languages speech sounds can be ranked on a *sonority hierarchy*, ranging from the least sonorous stop consonants to the most sonorous glides and vowels, with fricatives, nasals, and liquids having an intermediate, progressively more sonorous status. Sonority is related to the degree of openness of the vocal apparatus during speech (e.g., Goldsmith, 1990; MacNeilage, 1998; Selkirk, 1982) and hence to relative loudness, perceptibility and acoustic intensity (but see Harris, 2006). From this perspective, vowels would be processed differently from consonants because the former are more sonorous than the latter. Such a view also contends that the sonority of consonants affects how they are processed in relation to pitch in songs. In particular, the more sonorous nasals may be more apt to support pitch variations than the less sonorous stops, and hence their processing may be more integrated with melody. By contrast, if vowels and consonants were processed differently because they carry different functions in speech processing, no difference between nasals and stops should be found.

The aim of Experiment 4 was to examine the interactions between consonants and melody in sung nonwords using a *condensation test* (cf. Posner, 1964) in which no single dimension can serve as the relevant basis for classification. Finally, the filtering and redundancy tests of Experiment 5 were aimed at checking that the integrality between vowels and intervals did not result from acoustical interactions between the spectral characteristics of vowels and their pitch. To this end, we used a synthesized material in which these parameters were carefully controlled.

2. Experiment 1 – interactions between vowels and intervals in sung nonwords: filtering and redundancy tests

Participants had to classify bisyllabic nonwords sung on two-note intervals on the basis of the identity of either the nonword (*phonological task*) or the interval (*melodic task*), in the three conditions defined by Garner (1974). The nonwords differed from each other by the identity of their final vowel /ɔ̃/ vs. /ø/ and the intervals varied in their melodic contour, either ascending or descending (see Table 1).

Within each condition, the task remained formally the same: to associate each presented nonword (in the phonological task) or each interval (in the melodic task) to one of two response keys. In the baseline condition, only the dimension relevant for the task varied (e.g., for the phonological task, only the nonwords varied and the interval remained the same). In the redundant condition, there was a systematic correlation between the interval and the nonword variations. In the orthogonal condition, the four associations were presented and listeners had to ignore the irrelevant variations (see Table 2).

Based on their acoustical and functional characteristics, we hypothesized that vowels and melodic intervals would interact, leading to an integrality pattern, i.e., to facilitation in the redundant condition and interference in the orthogonal condition.

2.1. Method

2.1.1. Participants

Twenty-five undergraduate university students participated for course credits. One was discarded from further analyses because his error rate in the baseline of the melodic task was two-standard deviations above the mean. Of the 24 remaining participants, 20 were women; the mean age was 20.8 years (range: 18–36). Nineteen participants had never learned to play music, and five had no more than

Table 1
The different combinations of nonwords and intervals for the V1-material of Experiment 1. The F0 contour is marked in blue in the spectrograms.

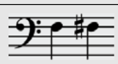
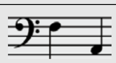
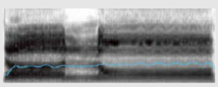
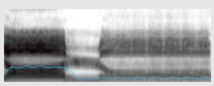
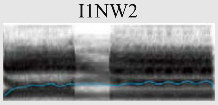
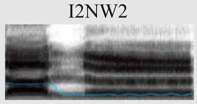
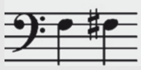
Experiment 1			
V1-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dalɔ̃/	I1NW1 	I2NW1 
Nonword 2 (NW2)	/dalø/	I1NW2 	I2NW2 

Table 2
Example of response attribution for the different combinations of tasks and conditions in the speeded classification of Experiment 1 for the V1-material. In each task (melodic and phonological), each participant was presented with one of the baseline materials in the baseline condition, with one of the redundant materials in the redundant condition and with the orthogonal condition.

Condition	Task: Melodic – Response x	Melodic – Response y	Phonological – Response x	Phonological – Response y
Baseline 1	 /dalɔ̃/	 /dalɔ̃/	 /dalɔ̃/	 /dalø/
or Baseline 2	 /dalø/	 /dalø/	 /dalɔ̃/	 /dalø/
Redundant 1	 /dalɔ̃/	 /dalø/	 /dalɔ̃/	 /dalø/
or Redundant 2	 /dalø/	 /dalɔ̃/	 /dalø/	 /dalɔ̃/
Orthogonal	 /dalɔ̃/	 /dalɔ̃/	 /dalɔ̃/	 /dalø/
	 /dalø/	 /dalø/	 /dalɔ̃/	 /dalø/

four years of music training that they had stopped at least five years ago.[4]

2.2. Material

Auditory examples of the stimuli used in all experiments can be heard on the website http://www.brams.umontreal.ca/plab/research/Stimuli/kolinsky_et_al/index.html.

Stimuli were CVCV bisyllabic pronounceable nonwords sung on two-note intervals. The slightly ascending F3-F3# and the descending F3-A2 minor intervals were combined with the nonwords /dalɔ̃/ and /dalø/ (henceforth, *V1-material*). They were sung by a professional baritone to avoid the major phoneme distortions linked to high frequencies generally observed in female opera singers (Scotto di Carlo, 2007a; Scotto di Carlo & Germain, 1985). They were recorded on mini-disc in an anechoic room, with tempo indicated by a silent luminous metronome. Since the first syllable lasted for around 500 ms, and the second for more than 1000 ms, tempo was set at 120 beats/min, with 1 pulsation for the first note and 2 pulsations for the second.

To avoid response strategies based on obvious physical correlations between, for example, pitch and quality, or length and intensity of the vowel, three physically different instances of each stimulus were selected among five recordings. The selected exemplars presented similar average duration of the first and second syllables, similar F0 and vibrato of the first syllable and accurate pitch of the second note. Care was taken to match durations of the first and second syllables across response categories, so that duration could not be a valid predictor of the response. The stimuli are shown in Table 1 and a thorough description of the selected exemplars is provided in the Appendix.

The selected stimuli were normalized for loudness and inserted in 2500 ms files. In order to facilitate response

[4] The error rate criterion and the musical expertise criterion were identical in all experiments.

time measurements across materials and conditions, silent intervals of different durations were inserted before and after the stimulus so that the transition between the two notes and syllables was centered at 1250 ms.

As illustrated in Table 2, which shows an example of the combinations of each task and condition, baseline blocks only included two (out of the four possible) stimuli that varied only on the target dimension, the other dimension being held constant. Thus, there were two different baseline pairs of stimuli for each task, each participant being presented with only one pair. The redundant blocks also included only two stimuli, but they varied on both dimensions in a correlated way: each value of the target dimension was systematically associated with one of the two values of the other dimension. There were thus also two different redundant pairs for each task, each participant being presented with only one pair. In the orthogonal condition, all four possible stimuli were presented. Each participant was presented with three blocks of trials corresponding to the three conditions. Each block included 72 trials, presented in random order (12 presentations of each exemplar of each of the two different stimuli in the baseline and redundant conditions; six presentations of each exemplar of each of the four different stimuli in the orthogonal condition).

2.3. Procedure

Participants were tested individually in a quiet room, with stimuli presented through headphones. Stimuli presentation and RTs were controlled via the *Psyscope* 1.2.5. PPC software (Cohen, Macwhinney, Flatt, & Provost, 1993) on a Macintosh G3 computer. Errors and RTs were recorded via a button box, using only the left- and rightmost keys and RTs were synchronized to stimulus onset.

Each participant performed the melodic and phonological tasks in the three conditions (baseline, redundant and orthogonal), leading to a total of six experimental blocks. Participants first performed the three conditions (blocks) in one task (melodic or phonological) followed by the other. Order of tasks and conditions was counterbalanced across participants, as well as key assignment and ascription of the specific baseline and redundant pairs.

Instructions were presented simultaneously on the screen and orally by the experimenter and emphasized both speed and accuracy. Participants were asked to classify the sung stimuli into two categories. In the melodic task, they were required to attend only to the melodies and to press one response key for ascending and the other for descending intervals. In the phonological task, they were told to attend only to the syllables and to press one response key for /dalɔ/ and the other for /dalø/. Table 2 shows an example of the combinations of each task and condition. In order to familiarize participants with the task, 12 practice trials preceded each experimental block. Auditory feedback was provided during the training (different beeps were triggered by correct, erroneous and timeout – longer than 2500 ms – responses); only the timeout feedback was provided during experimental blocks. The experiment took about 40 min.

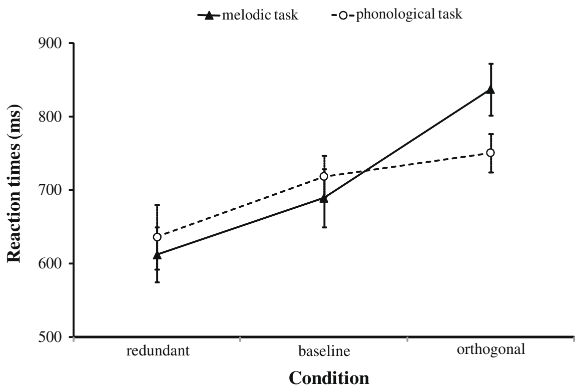


Fig. 1. Average RTs for correct classification responses to V1-material as a function of task (melodic: triangles and full lines; phonological: circles and dotted lines) and condition. Error-bars represent the standard-error of the mean.

2.4. Results and discussion

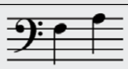
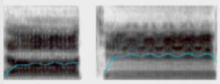
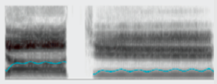
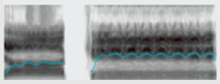
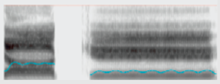
The error rate averaged over task, condition and participants was 6.4%. Further analyses were performed only on RTs to correct responses,[5] except for the analyses aimed at addressing potential discriminability differences between tasks. These also took accuracy into account, since the relative discriminability of the involved dimensions is critical for interpreting the pattern of dimensional interactions (Garner, 1974; Garner & Felfoldy, 1970). The accuracy analysis revealed a slight discriminability difference, with less errors in the phonological (2%) than in the melodic (9%) task, $t(23) = 2.3$, $p < .05$, without RT difference, $t(23) = -.89$, $p > .10$.

The analysis of variance (*ANOVA*) on RTs with task (melodic vs. phonological) and condition (redundant, baseline and orthogonal) as within-subject variables[6] showed no task difference, $F < 1$. The main effect of condition was significant, $F(2, 46) = 36.58$, $p < .001$: in comparison to baseline, there was interference in the orthogonal condition, $F(1, 23) = 37.86$, $p < .001$, and facilitation in the redundant condition, $F(1, 23) = 13.70$, $p < .001$. The interaction between task and condition was also significant, $F(2, 46) = 5.24$, $p < .01$. As illustrated in Fig. 1, interference was larger in the melodic (147 ms) than in the phonological (32 ms) task, $F(1, 23) = 8.93$, $p < .01$; the difference between the baseline and orthogonal conditions was highly significant in the melodic task, $F(1, 23) = 38.01$, $p < .001$ but only tended towards significance in the phonological task, $F(1, 23) = 3.087$, $p < .10$. The redundancy gain was significant in both tasks (phonological: 82 ms, $F(1, 23) = 7.51$, $p < .025$; melodic: 77 ms, $F(1, 23) = 7.30$, $p < .025$), and of similar amplitude, $F < 1$.

[5] In all analyses, these RTs were estimated by subtracting 1250 ms from the RTs measured from the beginning of the stimuli, as the crucial information for the classification task (the transition between the two notes and syllables) was centered at 1250 ms.

[6] In all analyses, in the first step the orders of task and condition were included as between-subjects variables. Since the effects and interactions including these variables were not significant, they were removed from the reported analyses for sake of clarity.

Table 3
The different combinations of nonword and interval for the stop-C-material of Experiment 2. The F0 contour is marked in blue in the spectrograms.

Experiment 2			
Stop-C-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/daty/	I1NW1 	I2NW1 
Nonword 2 (NW2)	/daky/	I1NW2 	I2NW2 

Interference in the orthogonal condition was thus modulated by task. This asymmetric interference can be explained by the slight difference of discriminability between dimensions observed in the error pattern. Indeed, if the values were more salient on one dimension than on the other, the dimension with more discriminable values would be easier to process and hence would have more opportunity to interfere on the processing of the dimension with less discriminable values (Ben Artzi & Marks, 1995; Garner & Felfoldy, 1970). However, the accuracy difference favoring nonwords cannot account for the almost significant interference effect observed in the phonological task. Variations on the more difficult dimension, the interval, tended nevertheless to interfere with the processing of the relatively easier dimension, the nonword. In addition, there was significant facilitation in the redundant condition, without between-task difference in the size of this effect. These results suggest that vowels and intervals behave like integral dimensions.

3. Experiment 2 – interactions between stop consonants and intervals in sung nonwords: filtering and redundancy tests

Our hypothesis that vowels and melodic intervals are integral dimensions was supported by the results of Experiment 1. In Experiment 2 we tested the additional hypothesis that consonants and intervals are less integrated. This can be either due to the acoustic properties of consonants, which prevent them from carrying melodic information, or to the different linguistic function and higher linguistic specificity of consonants compared to vowels.

3.1. Method

3.1.1. Participants

Thirty-six subjects who had not participated in Experiment 1 took part in Experiment 2. Two of them were discarded because of high error rates in the baseline condition and one because of musical expertise. The remaining 33 participants had an average age of 20.2 years (range: 18–23) and included 30 women. Most (24) had never learned music. They participated either for course credits (25 participants) or for financial compensation (eight participants).

3.2. Material and procedure

The stimuli, recorded by the same baritone in the same conditions as in Experiment 1, were the nonwords /daty/ and /daky/ varying on their middle, stop consonant sung on the major ascending interval F3-A3 and the major descending interval F3-C3 (henceforth, *stop-C-material*). As shown in the spectral representation of these stimuli in Table 3, the consonants did not provide any F0 information. Procedure was identical to the one of Experiment 1.

3.3. Results and discussion

Average error rate was 2.2%. In the baseline condition, a significant discriminability difference between the phonological and melodic task was found on RTs, $t(32) = 4.57$, $p < .001$, but not on accuracy, $t(32) = 1.38$, $p > .10$.

The ANOVA with task and condition as within-subject variables[7] showed a significant effect of task, $F(1, 32) = 54.18$, $p < .001$, favoring nonwords. Both the effect of condition, $F(2, 64) = 15.53$, $p < .001$, and the interaction between condition and task, $F(2, 64) = 7.55$, $p < .001$, were significant. As depicted in Fig. 2, the interaction arises from the effect of condition being restricted to the melodic task, with 101 ms facilitation in the redundant condition, $F(1, 32) = 34.68$,

[7] In a first analysis, the type of incentive received by the participants (course credits or payment) was an additional between-subjects variable. Because it had no significant effect and did not interact with either task or condition, $F < 1$, this variable was removed from further analyses.

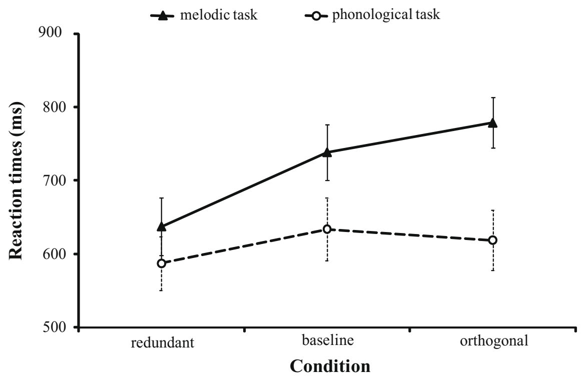


Fig. 2. Average RTs for correct classification responses to stop-C-material as a function of task (melodic: triangles and full lines; phonological: circles and dotted lines) and condition. Error-bars represent the standard-error of the mean.

$p < .001$, but no significant interference (40 ms), $F(1, 32) = 2.84$, $p > .10$. Conditions did not differ in the phonological task, $F(2, 64) = 1.84$, $p > .10$.

The redundancy gain restricted to the melodic task can easily be accounted for by the discriminability difference observed on RTs. In the redundant condition, participants could strategically rely on the dimension that is easier to discriminate (here, the nonwords), hence presenting facilitation in that condition in comparison to the hardest baseline (Maddox & Ashby, 1996). Shepp, Barrett, and Kolbet (1987) proposed an insightful test to detect this *selective serial processing strategy* (Garner, 1983): classification times are compared in the redundant and the faster baseline condition. If in the redundant condition participants based their judgment on the identity of the nonwords while ignoring the intervals, they should perform as in the baseline condition of the phonological task. Paired-samples *t*-test confirmed this view, $t(32) = .05$, $p > .10$ with only a 4 ms difference between conditions.

The discriminability difference might also increase interference in the melodic task, as in Experiment 1. However, this was not the case. Altogether, the evidence for selective serial processing (explaining the redundancy gain) and the absence of significant interference (in spite of discriminability differences) do not support the notion that stop consonants and intervals are integral dimensions. On the contrary, the present results suggest that stop consonants are processed independently of melodic intervals.

This idea is reinforced by direct comparison of the outcomes for vowels and stop consonants, demonstrated by ANOVAs run on the size of the effects. The amount of redundancy gain was computed as the RT difference between the baseline and redundant conditions. The amount of Garner interference was computed as the RT difference between the baseline and orthogonal conditions. Material (i.e., experiment) was a between-subjects variable and task was a within-subject variable. There was no difference in redundancy gain between materials (80 ms on vowels, 73 ms on consonants), $F < 1$, presumably because participants used a selective serial processing strategy in the redundant condition on stop consonants. As expected however, the vocalic or consonantal nature of the material significantly modulated the interference effect, $F(1, 55) = 15.82$, $p < .001$, with more interference for the V1-material (90 ms, on the average) than for the stop-C-material (13 ms). Thus, in regards to the Garner interference, the vocalic or consonantal nature of the material did influence the amount of integration with the melody, with stronger processing interactions between vowels and melodic intervals than consonants and melodic intervals.

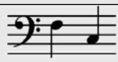
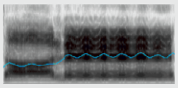
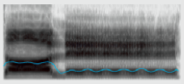
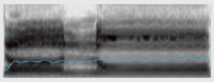
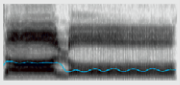
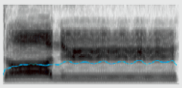
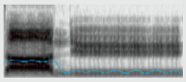
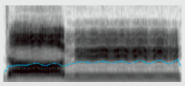
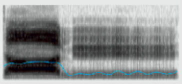
4. Experiment 3 – generalization to other vowels and to nasal consonants

Experiment 3 had three objectives. First, in Experiments 1 and 2, there was a systematic association between nonwords varying on vowels and minor intervals, and nonwords varying on consonants and major intervals. The size of these intervals also differed as the baseline discriminability had to be as similar as possible in each material for the phonological and the melodic tasks. This had the detrimental consequence of pairing the vocalic and consonantal nonwords with largely different interval pairs. Indeed, the V1-material was sung on either a slightly ascending or a largely descending interval, whereas the stop-C-material was sung on ascending and descending intervals of similar sizes. This variable might have contributed to the observed integration difference between vowels and consonants. Therefore, in Experiment 3, to control for the effect of interval size, new nonwords varying on vowels were sung on the intervals that were associated to the stop-C-material of Experiment 2, that is, F3-A3 and F3-C3. Similarly, new nonwords varying on consonants were sung on the intervals previously associated to the V1-material in Experiment 1: F3-#F3 and F3-A2.

Second, to generalize the results of Experiment 1 to new vowels, we chose another vocalic contrast: the nonwords /dalɛ/ and /dalɔ/ (henceforth, *V2-material,* see Table 4a). We expected to replicate the integrality pattern with these new vowel-interval associations.

Finally, to examine the contribution of the phonemes duration and sonority to their interaction with the melody, we used a contrast of nasal consonants, /n/ and /m/ (henceforth, *nasal-C-material,* see Table 4b). These nasals, which are among the most sonorous consonants in French, are continuous consonants that allow air to flow through the nose. Therefore, they are more likely to carry melodic information than the voiceless, discontinuous stop consonants used in Experiment 2. Inspection of the spectrogram of the stop-C-material (Table 3) and the nasal-C-material (Table 4b) confirms this view: whereas no pitch information is carried by the stop consonants (which seem to break the melodic line), the nasals support a continuous melodic contour. Furthermore, the transition between the first and the second tone are carried by these nasals, thereby contributing to the melodic contour. Two outcomes are possible with this nasal material. Given that integrality can be seen as a continuum (Grau & Kemler-Nelson, 1988), if for acoustical reasons vowels were more integrated with the melody than consonants, the result pattern for the nasal consonants should be intermediate between integrality

Table 4
The different combinations of nonword and interval for the V2-material (a) and the nasal-C-material (b) of Experiment 3. The F0 contour is marked in blue in the spectrograms.

Experiment 3			
a. V2-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dalɛ/	I1NW1 	I2NW1 
Nonword 2 (NW2)	/dalɔ/	I1NW2 	I2NW2 
b. Nasal-C-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dany/	**I1NW1** 	*I2NW1* 
Nonword 2 (NW2)	/damy/	*I1NW2* 	**I2NW2** 

and separability. If, on the contrary, the distinction between vowels and consonants were more categorical and rested on their different linguistic functions, results should be similar to the ones obtained in Experiment 2 for stop consonants.

4.1. Method

4.1.1. Participants

Fifty-seven subjects who had not participated in the previous experiments took part in Experiment 3. Nine were discarded: six because of poor performance in the baseline condition and three because due to musical training exceeding our criterion. Among the 48 remaining participants, 26 were women; mean age was 22.6 years (range: 18–54); 21 had never learned music. Twenty were unpaid volunteers, 13 were undergraduate students in psychology who earned course credits and 15 were university students paid for their participation.

4.2. Material

In the V2-material (Table 4a), the ascending F3-A3 and descending F3-C3 major intervals were combined with the nonwords /dalɛ/ and /dalɔ/. In the nasal-C-material (Table 4b), the F3-F3# and F3-A2 minor intervals were associated with the nonwords /dany/ and /damy/ varying by the place of articulation of the middle, nasal consonant. Stimulus recording and editing conditions were the same as in the previous experiments, except that stimuli were sung by a different professional baritone. The spectrograms of these stimuli are shown in Table 4.

4.3. Procedure

Procedure was the same as in the previous experiments, except that each participant was presented with both the V2- and the nasal-C-materials in one session. Order of presentation of the materials was counterbalanced between subjects, each participant completing the three conditions in the two tasks on one material before being presented with the other material. The experiment lasted for about 80 min.

4.4. Results and discussion

The speeded classification tasks were relatively easy: average error rates were 1.8% for the V2-material and 2.4% for the nasal-C-material. In the baseline condition, accuracy was similar in the phonological and melodic tasks for the V2-material: $t(47) = -1.4$, $p > .10$. Moreover, the 20 ms trend of performing the phonological task faster than the melodic task did not reach significance, $t(47) = 1.79$, $p < .08$. For the nasal-C-material, no discriminability difference between the phonological and the melodic dimensions was observed, either on accuracy scores, or on RTs: $t(47) = -.017$ and $= 1.15$, respectively, $p > .10$ in both cases.

The ANOVA took material (V2- vs. nasal-C-material), task (phonological vs. melodic) and condition (baseline, redundant, or orthogonal) as within-subject variables,[7] and revealed significant main effects of all three, $F(1, 47) = 14.29$, 20.94, and 105.74, respectively, all $ps < .001$. RTs were 30 ms shorter on average for the V2- than for the nasal-C-material, and performance was better in the phonological than in the melodic task. Overall, the effect of condition corresponded to an integrality pattern. However, condition interacted with material, $F(2, 94) = 12.7$, $p < .001$, and with task, $F(2, 94) = 18.68$, $p < .001$. The interaction between material and task was also significant, $F(1, 47) = 6.61$, $p < .025$, as was the second-order interaction between material, condition and task, $F(2, 94) = 17.74$, $p < .001$.

The decomposition of this second-order interaction shows significant effects of task, $F(1, 47) = 30.64$, $p < .001$, and condition, $F(2, 94) = 107.48$, $p < .001$) for the V2-material. In comparison to baseline, RTs were 32 ms faster in the redundant condition, $F(1, 47) = 20.81$, $p < .001$, and were slower in the orthogonal condition, $F(1, 47) = 115.84$, $p < .001$. The interaction between task and condition was also highly significant for the V2-material, $F(2, 94) = 33.78$, $p < .001$. Although interference in the orthogonal condition was significant in both the melodic and phonological tasks – as can be seen in Fig. 3a – with $F(1, 47) = 108.27$, $p < .001$ and $F(1, 47) = 10.26$, $p < .005$, respectively, this effect was far more pronounced when participants processed the melodic (130 ms) than the phonological (22 ms) contrast, $F(1, 47) = 59.76$, $p < .001$. As in Experiment 1, this asymmetric interference could be explained by the slight difference in discriminability between the two dimensions.

For the nasal-C-material, the effect of task was significant, $F(1, 47) = 5.20$, $p < .05$, but less prominent than in the V2-material, thereby explaining the interaction between task and material. The effect of condition was also significant, $F(2, 94) = 23.45$, $p < .001$: in comparison to baseline, there was a significant 42 ms interference effect in the orthogonal condition, $F(1, 47) = 28.32$, $p < .001$, but no significant facilitation (11 ms) in the redundant condition, $F(1, 47) = 1.82$, $p > .10$ (see Fig. 3b). Task and condition did not interact, $F < 1$.

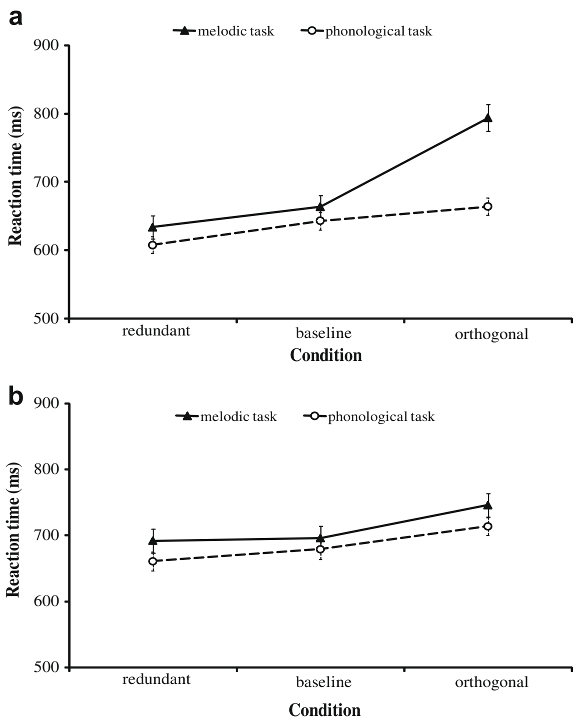


Fig. 3. Average RTs for correct classification responses to V2-material (a) and nasal-C-material (b) as a function of task (melodic: triangles and full lines; phonological: circles and dotted lines) and condition. Error-bars represent the standard-error of the mean.

Thus with vowels, as in Experiment 1, there was not only mutual (although asymmetric) interference, but crucially, mutual and symmetric facilitation in the redundant condition. Because the intervals of the V2-material were identical to those of the stop-C-material of Experiment 2, the specific intervals used do not seem to play a fundamental role in the different result patterns found for vowels and consonants. The present results thus allow for a generalized conclusion that vowels and intervals are integral dimensions. By contrast, with nasal consonants, there was interference, but no significant redundancy gain.

The processing difference between the V2- and the nasal-C- materials was further supported by additional cross-material analyses. Indeed, the ANOVAs on interference and redundancy gain, with task and material as within-subject variables, showed that material did significantly modulate the amount of interference, $F(1, 47) = 10.03$, $p < .005$, with more interference for the V2- (76 ms) than for the nasal-C-material (42 ms). Concerning the redundancy gain, there was a trend towards larger facilitation for the V2- (32 ms) than for the nasal-C-material (12 ms), $F(1, 47) = 3.14$, $p < .09$. Thus, these analyses showing weaker interference and slightly weaker redundancy gain for the nasals compared to the V2-material confirm that consonants are less integrated than vowels with the processing of intervals.

Still, sonority of the consonants could contribute to their interactions with melody. To examine this question, we ran cross-experiment analyses to compare the interference effects and redundancy gains obtained with the nasals used here and with the stops of Experiment 2. If the lack of sonority and resulting inability to carry melodic information explained at least partly the separability between consonants and intervals, we should observe less facilitation and interference for the stops than for the nasals. This was not the case. The ANOVAs on the size of the effects, with task as within-subject variable and type of consonantal contrast as between-subjects variable, showed that consonantal contrast had a significant effect on the redundancy gain, $F(1, 79) = 12.10$, $p < .001$. However the direction of the effect was opposite to the one expected on the basis of sonority, with larger facilitation for stops (80 ms) than for nasals (12 ms). Moreover, the significant interaction between task and type of consonantal contrast, $F(1, 79) = 4.81$, $p < .05$, showed that the redundancy gain was larger for stops only in the melodic task, $F(1, 79) = 20.84$, $p < .001$, reflecting the selective serial processing strategy that participants had adopted for this task in Experiment 2. Regarding Garner interference, neither the effect of type of consonantal contrast, nor the interaction of this variable with task, reached significance, $F(1, 79) = 2.51$ and 1.845, both $ps > .10$. The present results thus provide no evidence for the contribution of sonority to the interaction between lyrics and melody: the more sonorous nasal consonants were not more integrated with interval processing than the least sonorous, stop consonants.

In sum, these results allow generalization of results previously observed in Experiments 1 and 2. That is, the weaker interference and slightly weaker redundancy gain observed for nasals compared to vowels, suggests less integration of consonants with the processing of intervals, whatever the sonority class of consonants.

However, nasals did elicit some interference, an outcome that should be clarified since significant interference without facilitation is not a typical separability pattern. It may either reflect listeners' inability to selectively focus attention on the target dimension despite genuine dimensional separability (Thibaut & Gelaes, 2002) or the presence of an *emergent feature* (Pomerantz & Garner, 1973). The aim of Experiment 4 was to disentangle these two interpretations.

5. Experiment 4 – condensation test

The occurrence of an interference effect without redundancy gain for the nasals used in Experiment 3 corresponds neither to an integrality nor to a separability pattern. According to some authors, interference without facilitation merely arises from difficulties in attending selectively to separate dimensions, because of task difficulty and/or lack of discriminability between the values of the dimensions. This has been shown, for example, in developmental studies on attentional filtering capacities (e.g., Thibaut & Gelaes, 2002).

Alternatively, interference without facilitation may suggest *configural interaction* between dimensions, due to the presence of an *emergent feature* (Garner, 1974; see also Pomerantz & Garner, 1973). In the visual domain, interference without facilitation has been observed when participants have to classify pairs of parentheses according to, say, the orientation of the left parenthesis, thus pooling ((and () on the one hand, and)) and) (on the other hand. Difficulty to selectively pay attention to the orientation of the left parenthesis would be linked to the emergence of new salient features of parallelism, symmetry and closure distinguishing ((and)) from () and) (. These emergent features are more salient and discriminable from one another than are the underlying dimensions (the individual parentheses), leading to a *configural superiority effec*t (Pomerantz, Sager, & Stoever, 1977) in discriminating () from)) in comparison to discriminating (and). In the filtering test, emergent features are not useful (they are not mapped onto response categories in a one-to-one fashion, cf. Garner, 1974, 1978b; Pomerantz & Garner, 1973; Pomerantz & Schwaitzberg, 1975), explaining a performance drop.

Although it is not obvious that a new dimension could emerge from the association between specific intervals and nonwords as those used in Experiment 3, we tested this alternative account of our results for the nasal-C-material by using a condensation test (also called *biconditional classification* by Garner, 1974). Condensation is a test in which no single dimension can serve as the relevant basis for classification: participants have to classify in the same category stimuli differing on the values of both dimensions. Considering the example above, it would require pooling ((and)) on the one hand, and () and) (on the other hand. As illustrated, emergent features now distinguish these two response categories, and hence condensation is carried out more easily than filtering (in which these features are not mapped onto response categories). In contrast, in the absence of an emergent feature, condensation leads to poorer performance than filtering, in particular when dimensions are separable (e.g., Fitts & Biederman, 1965; Gottwald & Garner, 1972, 1975; Keele, 1970; Morin, Forrin, & Archer, 1961). Indeed, in this case the only way to perform correctly is to pay attention to both dimensions at the same time, resulting in slower RTs than when filtering only requires consideration of a single dimension.

Based on this logic, we presented the nasal-C material used in Experiment 3 for a condensation test that required participants to pool stimuli differing on the values of both dimensions (interval and nonword) into the same response category. Indeed, the two redundant subsets (of two stimuli each) had to be sorted into a single class. For example, as illustrated by the font styles in Table 4b (bold vs. non bold), participants had to pool /damy/ sung on F3-F3# (I1NW2) and /dany/ sung on F3-A2 (I2NW1) into the same response category, and /damy/ sung on F3-A2 (I2NW2) and /dany/ sung on F3-F3# (I1NW1), into the other response category. If in the nasal material new features emerged from the association between consonantal and interval values, this test should be easier than the filtering test of Experiment 3.

5.1. Method

5.1.1. Participants

Twenty-four undergraduate students who had not participated in the previous experiments took part in Experiment 4 for course credit. There were 20 women; mean age was 20.7 years (range: 17–31); 12 participants had never learned music.

5.2. Material and procedure

We presented the four stimuli, with three different exemplars of each stimulus, from the orthogonal condition of the nasal-C-material of Experiment 3. As illustrated by the font styles in Table 4b, the task required the two redundant subsets (of two stimuli each) to be sorted into a single class (both bold stimuli together vs. both non bold stimuli together). Participants had to pool stimuli differing on the values of both dimensions – interval and nonword – into the same response category.

The task included one experimental block of 72 trials (18 presentations of each of the four stimuli, with six presentations of each of the three exemplars for one stimulus), preceded by 12 practice trials (three presentations of each of the four stimuli, with three different exemplars for each stimulus). Order of trials within each block was random and response assignment to keys was counterbalanced across participants. The procedure was the same as in previous experiments, with testing time reduced to about 10 min.

5.3. Results and discussion

Confirming the notion that condensation is a fairly difficult task, delayed responses (longer than 2500 ms and hence discarded from further analyses) were numerous, reaching about 8% of the trials. On the remaining trials, error rates were also quite high: 28%, on average.

We compared classification times as well as accuracy rates between the condensation test and the filtering test (i.e., the orthogonal condition) of Experiment 3, separately for each dimension. Participants were both slower and less accurate in condensation than in filtering (see Fig. 4a and b, respectively). This holds true for the comparison of condensation with both the melodic filtering task, with differences of 271 ms for RTs and 23% for errors, $t(70) = 9.81$ and 9.26, $p < .001$ in both cases, and the phonological filtering task, with differences of 304 ms for RTs and 25% for errors $t(70) = 13.90$ and 11.04, $p < .001$ in both cases.

These results suggest that no new feature emerged from the association between specific intervals and nonwords in the nasal material. Had it been the case, performance would have been better in the condensation than in the filtering test (Garner, 1978b). Since the reverse pattern was observed, the interference effect reported in Experiment 3 for this material was likely related to a lack of selective attention to otherwise separable dimensions. Thus, consonants seem to be more separable from melody than vowels. Experiment 5 was aimed at examining the origin of the repeated observation of integrality between vowels and intervals.

6. Experiment 5 – filtering and redundancy tests on synthesized vocalic material

In Experiments 1 and 3, we observed an integrality pattern for the vocalic materials, which was not the case for the consonantal materials of Experiments 2 and 3. Such a difference suggests that vowels and intervals are at least partly processed by common auditory mechanisms, in contrast to consonants and intervals.

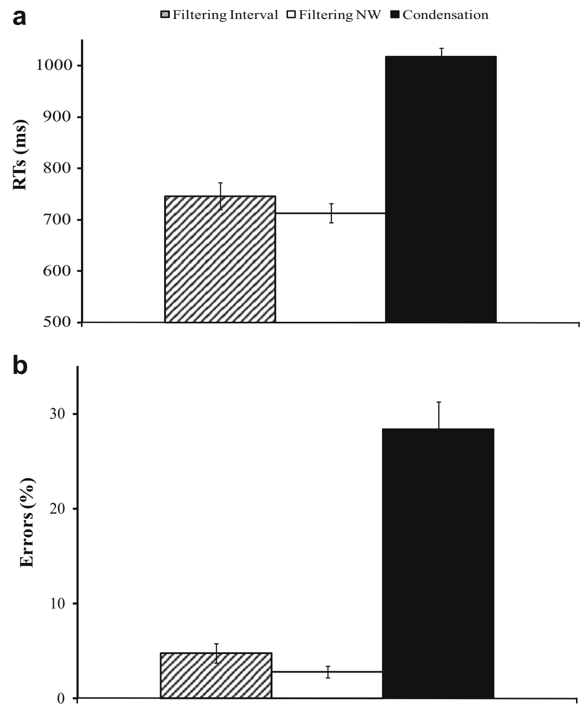
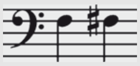


Fig. 4. Average RTs (a) and percentage of errors (b) for the nasal-C-material in the filtering (Experiment 3) and condensation (Experiment 4) tasks; NW = nonwords. Error bars represent the standard-error of the mean.

Alternatively, this response pattern may reflect physical interactions between the linguistic and musical dimensions. In singing, different pitches can alter the intelligibility of speech. This is well known by singers and by opera lovers with studies on song production and intelligibility having confirmed the difficulty of understanding sung words (for a review, see Scotto di Carlo, 2007a, 2007b). This effect seems to depend mainly on the intelligibility of vowels, which is inversely proportional to their pitch (e.g., Gregg & Scherer, 2006; Scotto di Carlo & Germain, 1985; Sundberg, 1982): in songs, the spectral characteristics of the vowels vary with pitch height (for a review, see Astesano, Schön, & Besson, 2004).

To control for such interactions in the V-materials, we examined the spectrum of the second vowel (Boersma & Weenink, 2007). The frequencies of the first (F1) and second (F2) formants of the vowels of the V-materials of Experiments 1 and 3 (averaged over the three exemplars of each stimulus) are shown in Table 5 (the detailed values for each exemplar are provided in the Appendix). The difference in pitch (A2 and #F3) slightly influenced the formant frequencies of the vowel /ø/ of the V1-material (Experiment 1), especially by decreasing the frequency of F2. In the V2-material (Experiment 3), the effect of pitch change on the formant frequencies was more systematic, probably because the average F0 was higher than in the V1-material. The increase in pitch from C3 to A3 tends to slightly decrease F1 for /dalɔ/ and to increase it for /dalɛ/. By contrast, pitch rise increases F2 for /dalɔ/ and decreases it for /dalɛ/. Thus, these vowels sung on high pitches

Table 5
Averaged values of F1 and F2 in the final vowel of the natural stimuli of the V-materials used in Experiments 1 (V1-material) and 3 (V2-material), in Hz.

V1-Material	Averaged F1 of the second vowel (Hz)	Averaged F2 of the second vowel (Hz)
Stimulus		
Averaged /dalɔ̃/ 	578	2139
Averaged /dalɔ̃/ 	583	2134
Difference	5	-5
Averaged /dalø/ 	403	1299
Averaged /dalø/ 	392	1246
Difference	-11	-53

V2-Material	Averaged F1 of the second vowel (Hz)	Averaged F2 of the second vowel (Hz)
Stimulus		
Averaged /dalɛ/ 	497	1706
Averaged /dalɛ/ 	511	1636
Difference	14	-70
Averaged /dalɔ/ 	545	838
Averaged /dalɔ/ 	520	893
Difference	-25	55

should be slightly less discriminable, as is usually the case in song production (Astesano et al., 2004; Scotto di Carlo, 2007a, 2007b).

The correlation between the spectral characteristics of the vowels and pitch height may have reinforced the integrality effect, at least for the V2-material. This could, for

Table 6
First and second formant frequency implemented in the source-filter synthesizer while programming (left) and measured afterwards in the synthesized stimuli (right) of Experiment 5.

Synthesized material Stimulus	Implemented F1 of the second vowel (Hz)	Implemented F2 of the second vowel (Hz)	Averaged measured F1 of the second vowel (Hz)	Averaged measured F2 of the second vowel (Hz)
/dalɛ/ 	600	1770	609	1778
/dalɛ/ 	600	1770	607	1780
Difference	0	0	-2	2
/dalɔ/ 	500	900	507	897
/dalɔ/ 	500	900	503	896
Difference	0	0	-4	-1

example, explain that it was only for this material that the redundancy gain and the Garner interference were significant in both the phonological and the melodic tasks. Although the spectral differences were variable from one item to the other (see Appendix), listeners may have used the fine acoustical differences between a high-pitched /ɛ/ and a low-pitched /ɛ/ to perform the task. In order to check for this possibility, a new V2-material was synthesized that did not include any acoustical correlation between pitch height and vowel quality.

6.1. Method

6.1.1. Participants

Twenty-six paid undergraduate students who had not participated in the former experiments took part in Experiment 5: eight were women, mean age was 25.8 years (range: 18–37), and nine had never learned music.

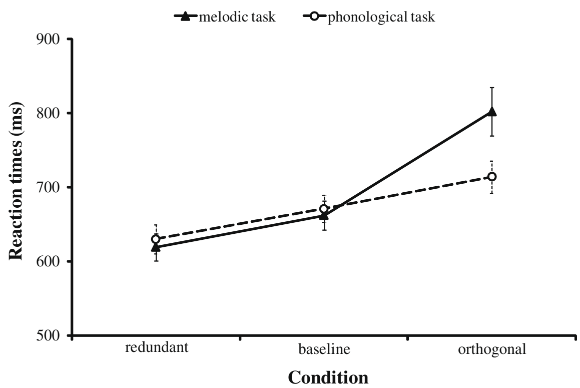


Fig. 5. Average RTs for correct classification responses with the synthesized V-material as a function of task (melodic: triangles and full lines; phonological: circles and dotted lines) and condition. Error-bars represent the standard-error of the mean.

6.2. Material

A source-filter synthesizer simulating the vocal tract was used to synthesize the four stimuli, namely the nonwords /dalɛ/ and /dalɔ/ sung on F3-A3 and F3-C3. For each vowel, five formants were modeled. Frequency of the first four formants determined the nature of the vowel and F5 was always 1000 Hz higher than F4. The central frequency values of the first four formants were fixed, so that they could not be influenced by pitch (Table 6, left panel, for F1 and F2 values). The stop consonant /d/ and the liquid /l/ were modeled by manipulating the formant transitions. The model also fixed F0 at 220 Hz for A3 and 130 Hz for C3. Syllable duration was 500 ms for the first syllable and 1330 ms for the second. In order to obtain a singing-like result, a vibrato deviating by 2% of F0 was added. Slight random variations of pitch, harmonics, duration and vibrato were also included to produce more natural singing. Because of these random variations, formant frequency was not identical across stimuli. However, comparison of the left and right panels of Table 6 shows that these differences were minimal, and importantly, that pitch differences did not modify F1 and F2 frequencies in a systematic way. As these random variations were also present in the first syllable and could influence classification judgments, the two syllables of each of the four stimuli were cross-spliced just before /l/, yielding four different exemplars of each stimulus.

6.3. Procedure

The procedural aspects were the same as in Experiments 1 and 2, except that, in order to include the 16 physically different stimuli (four exemplars of the four stimuli), there were 16 instead of 12 training trials. Similarly, experimental blocks counted 80 instead of 72 trials. The experiment lasted for about 40 min.

6.4. Results and discussion

Average error rate was 2.6%. There was no significant discriminability difference between the melodic and the phonological dimensions in the baseline condition, $t(47) = -.598$ on RTs, and = .778 on errors, both $ps > .10$.

The ANOVA run on RTs, with condition and task as within-subject variables, showed no significant effect of task, $F(1, 23) = 1.67$, $p > .10$. In contrast, both the effect of condition, $F(2, 46) = 43.89$, $p < .001$, and the interaction between task and condition, $F(2, 46) = 7.97$, $p < .001$, were significant. The effect of condition revealed interference in the orthogonal condition, $F(1, 23) = 40.99$, $p < .001$, and facilitation in the redundant condition, $F(1, 23) = 13.32$, $p < .001$, as compared to baseline (see Fig. 5). These results replicate those obtained in Experiment 3. As with natural singing, there was more interference in the orthogonal condition for the melodic than for the phonological task, $F(1, 23) = 8.93$, $p < .01$, although interference was significant in both tasks, with 140 and 43 ms costs respectively ($F(1, 23) = 29.70$, $p < .001$, and = 10.07, $p < .005$). Facilitation in the redundant condition did not differ as a function of task (43 and 41 ms, respectively, $F < 1$). In brief, these results correspond to a typical integrality pattern, except for asymmetric interference.

To check whether the natural vs. synthesized nature of the material influenced the pattern of interactions between vowels and melody, a further ANOVA was run on the two sets of V-materials used in Experiments 3 and 5, with type of singing (natural vs. synthesized) as a between-subjects variable and task and condition as within-subject variables. Crucially for the present purpose, the effect of type of singing was not significant, $F < 1$, and did not interact with condition, $F(2, 140) = 1.57$, $p > .10$. The second-order interaction between task, condition and type of singing was also not significant, $F < 1$. In brief, results with natural and synthesized materials were very similar. This suggests that the integrality pattern, revealed by the occurrence of a redundancy gain in addition to interference for vowels in both Experiments 1 and 3, reflects genuine psychological interactions and is not merely due to the acoustical correlations between pitch and vowel quality.

7. General discussion

In the present study, we examined whether the phonological and melodic dimensions of sung material are processed independently or in an integrated way. For this we used the filtering (Experiments 1, 2, 3 and 5) and condensation (Experiment 4) tests designed by Garner (1974) with auditory CVCV nonsense sung syllables. Moreover, we compared materials with varying vowels (V-materials: Experiments 1, 3, and 5) and varying consonants, consisting of either stops (stop-C-material: Experiment 2) or nasals (nasal-C-material: Experiments 3 and 4).

The underlying motivation for this separate manipulation of vowels and consonants is two-fold. First, their physical characteristics are quite different, and vowels may constitute a more natural melodic support than the transient consonants. Second, both their phylogenesis and informative values suggest that they pertain to distinct processing systems, an assumption supported by the occurrence of double dissociations between consonant and vowel production (e.g., Caramazza et al., 2000). We therefore hypothesized that consonants and vowels bear different relationships with other human auditory processing, involving music perception. Song is an obvious case for this inquiry, since it involves both music and speech.

The outcomes of the present series of Experiments are summarized in Table 7. Overall, the results support the notion that in songs, consonants are more separable from melodic processing than vowels are. As a matter of fact, all experiments involving materials varying on vowels (Experiments 1, 3 and 5) revealed a clear integrality

Table 7
Summary of the main results of the five experiments.

	Experiment	Material(s)	Classification test	RTs in the melodic (M) and phonological (P) tasks	Result pattern and interpretation
VOWELS	Experiment 1	V1	Filtering and redundancy	M: Redundant < Baseline < Orthogonal P : Redundant < Baseline ⩽ Orthogonal	Redundancy gain and (asymmetric) Garner interference: Integrality
	Experiment 3	V2	Filtering and redundancy	M: Redundant < Baseline < Orthogonal P : Redundant < Baseline < Orthogonal	Redundancy gain and (asymetric) Garner interference: Integrality
	Experiment 5	Synthesized V2	Filtering and redundancy	M: Redundant < Baseline < Orthogonal P : Redundant < Baseline < Orthogonal	Redundancy gain and (asymmetric) Garner interference: Integrality
CONSONANTS	Experiment 2	Stop-C	Filtering and redundancy	M: Redundant < Baseline = Orthogonal P: Redundant = Baseline = Orthogonal	No interference and redundancy gain compatible with serial processing strategy : Separability
	Experiment 3	Nasal-C	Filtering and redundancy	M: Redundant = Baseline < Orthogonal P : Redundant = Baseline = Orthogonal	No redundancy gain and Garner interference reflecting either emergent features or selective attention difficulties
	Experiment 4	Nasal-C	Condensation	Condensation > M and P Filtering (Exp 3)	No emerging dimension : Separability

pattern, with both interference in the orthogonal condition and facilitation in the redundant condition. This was not the case in experiments involving materials varying on consonants (Experiments 2 and 3).

For vowels, it is worth noting that the integrality pattern occurred in both the phonological and melodic tasks, and for both natural and synthesized sung stimuli. All acoustic correlations between the two dimensions were eliminated from the synthesized stimuli of Experiment 5, so that the observed results likely reflect the perceptual processing of songs. For both naturally sung and synthesized V-materials, the participants' classifications fit an Euclidean metric of (dis)similarity, with perceptual distance between the stimuli being enhanced in the redundant condition. Such a response pattern shows that the underlying dimensions are processed in a somewhat integrated, unitary fashion. Furthermore, the integrality between vowels and intervals has been replicated with different melodic intervals (Experiments 1 and 3) and thus is largely independent of the identity of the intervals.

Although interference reached significance in both the melodic and the phonological tasks (except in the phonological task of Experiment 1), it was asymmetric in all experiments on vowels, with stronger interference from the vocalic variations in the melodic task than from the melodic variations in the phonological task. This asymmetry may result from the higher discriminability of the linguistic information compared to the musical information. This is a classical outcome in the literature on song processing (Peretz, Radeau, & Arguin, 2004; Serafine, Crowder, & Repp, 1984), especially with musically untrained participants. Still, this explanation does not seem to account for the asymmetry found with the synthesized material of Experiment 5, in which there was no discriminability difference at all between the target dimensions. Hence, asymmetrical interference may reveal more fundamental processing differences between vowels and intervals, such as the level at which the dimensions interact. As suggested by Melara and Marks (1990), asymmetries in Garner interference can reflect the fact that the least interfering dimension, here the interval, is processed later than the most interfering dimension, here the vowel. But the interpretation that phonetic processing occurs earlier than a supposedly more acoustical, pitch-related processing seems counterintuitive, especially in view of results showing the opposite asymmetry for stable pitch levels and consonants (Wood, 1974, 1975). Still, melodic intervals are probably musically more complex than pitch height and may recruit music-specific mechanisms (Peretz & Coltheart, 2003). For example, whereas musical tones in isolation activate spatial mental representations similar to the ones associated to many ordered sequences, intervals do not evoke these general spatial representations (Lidji, Kolinsky, Lochy, & Morais, 2007). Also, patients suffering from acquired amusia are able to discriminate isolated pitches but exhibit deficits for discriminating the same tones when these are embedded in a melody (Peretz & Kolinsky, 1993; Peretz et al., 1994). In this perspective, processing of melodic intervals and contour rests on sophisticated musical processes that take place later than the phonetic processing of vowel contrasts. In summary, vowel processing seems to interact with but precede interval processing in nonmusicians. This asymmetry would be specific to a "musical mode of processing", given that vowels and tonal contours induce symmetric interference in tonal languages (Lee & Nusbaum, 1993; Repp & Lin, 1990).

In contrast with the results on vowels, the resulting pattern for sung nonwords that differed by their consonants never corresponded to typical integrality. For stop consonants (Experiment 2), there was no interference in the orthogonal condition but a significant redundancy gain that was limited to the melodic task. This could be easily explained by differences in discriminability between the nonwords and intervals, leading participants to strategically attend to the more salient dimension to perform the redundancy test. This serial processing strategy (Garner, 1974) has been statistically confirmed in the present data. Thus, stop voiceless consonants and intervals seem to be separable and not integral dimensions.

The separability of stop consonants and melody is consistent with the detrimental effect of consonants on musicality reported in studies of song production (McCrean & Morris, 2005; Scotto di Carlo, 1993; Sundberg, 1982). As can be seen in Table 4b, voiceless stop consonants consist of a silence followed by a noise-burst, making such consonants incompatible with melody continuity. In order to assess whether separability between consonants and intervals was solely due to these acoustical constraints, we used nasal consonants in Experiment 3.

Indeed, patterns of interactivity may reflect a continuum of separable-to-integral processing rather than two categorically distinct processing modes (Garner, 1974; Grau & Kemler-Nelson, 1988). Phonemes themselves vary continuously in sonority, and the more sonorous consonants may be more apt to support pitch variations than the less sonorous ones, and hence may be more integrated with melody. In Experiment 3, we examined whether sonority modulates the interactivity pattern by using more sonorous consonants, namely nasals. With this material, there was no redundancy gain and only interference in the orthogonal condition of the melodic task. The notion that this interference reflects a failure of selective attention to the target dimension (Thibaut & Gelaes, 2002), rather than the processing of a new feature that would have emerged from interaction between the underlying dimensions (e.g., Garner, 1974), was confirmed in Experiment 4. Indeed, with the same nasal material, condensation was far more difficult than filtering. This evidence of separability between nasal consonants and intervals, added to the fact that the pattern of interference did not differ significantly between the stop and the nasal materials, suggests that the sonority of the segments is not the key to their interactions with melody.

To summarize, vowels seem to merge, or at least to interact with intervals during song processing, but consonants do not. In addition, our results show that the close relation between vocalic and melodic variations has processing consequences even when all acoustic correlations between vowel quality and pitch height are eliminated, as in the synthesized material used in Experiment 5, and that the separability of consonants and melodic variations cannot be explained solely by their lower sonority level, in comparison to vowels. Thus, the results observed here must reflect a higher processing level than the acoustical one.

Of course, the kind of interval information provided by vowels and sonorous (nasal) consonants in the present sung materials is different (as is the case in natural singing), and this may at least in part explain the discrepancy in processing interactions with melody. Inspection of the pitch trajectories in the stimuli (Tables 1–4) indicates that vowels are sung on relatively stable pitch levels (if one excludes vibrato), whereas nasals are located at the transition between two tones. That is, pitch changes occur at the same time as the consonant is articulated. Moreover, the consonants were far shorter in duration than the vowels, a phenomenon also typical of spontaneous singing (McCrean & Morris, 2005; Scotto di Carlo, 1993). A control for these two factors would be to use filtering and redundancy tests on contrasts made out of nasal consonants sung on a stable pitch, and on the same contrasts sung on melodic intervals, a kind of singing called humming. The duration of these nasals could also be manipulated. If the nasals hummed on stable pitch, but not those hummed on intervals, led to integrality, this would support the idea that the kind of musical information usually coupled to either vowels or consonants is crucial for processing interactions.

However, the differences in processing interactions of vowels and consonants with melody may also be related to other factors than these acoustic features. More generally, domain-specificity (or non-specificity), be it in the language or music domain, may rest on the representation of abstract properties. For example, in speech it has been shown that left-hemisphere structures mediate abstract properties of language rather than acoustic features. Neuroimaging studies of linguistic pitch processing in tonal languages, such as Thai or Chinese, have demonstrated that left-hemisphere structures are recruited for processing pitch contours in speakers of such languages, whereas non-speakers process identical stimuli via right-hemisphere mechanisms (e.g., Gandour, Wong, & Hutchins, 1998; Gandour et al., 2000; Klein, Zatorre, Milner, & Zhao, 2001).

In the present case, the linguistic function of vowels might be closer to the functions of pitch in music than the linguistic function of consonants. According to the studies reviewed in the Introduction, this seems actually a plausible view. In addition to being responsible for prosody, the information carried by vowels gives important information about syntactic structure, as well as about some aspects of the morphological system. In contrast, consonants play a minor role in signaling syntax (limited to juncture phenomena that signal constituency, Nespor & Vogel, 1986; Selkirk, 1984), but a major role in making lexical distinctions. Semitic languages provide an extreme illustrative case of this role, since in these languages lexical roots are formed exclusively by consonants, whereas vowels are inserted to indicate morphological patterns (McCarthy, 1985). In line with our results, Nazzi (2005) and Nazzi and New (2007) reported evidence that this dissociation of function is not related to the sonority of the phonemes. While exhibiting difficulties in learning new words differing by one vowel, infants were able to perform this task when the distinctive phoneme was a consonant, regardless of the sonority of the phoneme. Their performance was identical when the consonant was a stop as when it was a continuous, nasal, liquid or fricative consonant. Similarly, sonority *per se* seems unable to account for the processing specificity of consonants and vowels: it cannot explain the errors made by patients selectively impaired in either consonant or vowel production (Caramazza et al., 2000). Thus, vowels and consonants may carry specific kinds of information and serve distinctive linguistic functions, independently of their sonority characteristics. As a result, these phoneme classes would be processed differently, and would have different functional links with other domains like music.

In particular, it is clear that both music and language display a generative structure that yields an infinite number of possible experiences (Bharucha, Curtis, & Paroo, 2006). Coherently, both are rule-based. Thus, vowels and intervals may share an important syntactic and grammatical role within speech and musical systems, respectively. In contrast, along with their transient acoustical nature, consonants seem to have a more lexical function. Finding a parallel function in music may be harder; actually, the notion that music has its own semantics is still under debate (e.g., Bharucha et al., 2006; Koelsch et al., 2004).

In any case, the present results imply that neither the speech nor the musical system is homogeneously modular. Indeed, while many data point to domain-specificity in both speech and music processing (for overviews, see Patel, 2008; Peretz, 2006), the observed interaction between vowel and interval processing entails that speech and music processing are not totally independent. The difference in processing interactions of vowels and consonants with melody shows that, whatever the functional domain, modularity and interactivity may best be evaluated when one proceeds to a detailed level of empirical test, and are then likely to both be rejected (Peretz & Coltheart, 2003).

Future research is needed however to assess whether the processing interaction between the two systems depends on the actual functions that speech and music parameters play in a specific setting. For example, it has recently been suggested that learning a new language, especially in the first learning phase wherein one needs to segment new words, may largely benefit from the structuring properties of music in song. Compared to speech sequences, a consistent mapping in sung sequences of linguistic and musical statistical dependencies - the location of dips in transitional probabilities between adjacent syllables and/or tones (cf. Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996) – enhanced learning of an artificial language (Schön et al., 2008). Independent work in speech has shown that consonants are much more suitable than vowels to parse streams into words, using statistical dependencies, with "consonant words" significantly preferred over "vowel words" (Bonatti et al., 2005; Mehler et al., 2006). The present results suggest stronger processing interactions between vocalic and melodic variations than between consonantal and melodic variations. However it may be the case that lexically parsing sung streams depends more on the statistical dependencies between consonants than on the statistical dependencies between vowels. In other words, processing sung sequences through a *lexically-oriented* learning mechanism might induce other interactions between language and music in sung material than the ones observed here. This would also support the idea that studying the interactions between language and music at various and detailed levels of processing is the key for a better understanding of the similarities and specificities of each of these domains. If, as the expression goes, "it ain't over till the fat lady sings", then in this research area, she has not yet sung.

Acknowledgements

This research was supported by a grant of the *Human Frontier of Science Program* (RGP 53/2002, "An interdisciplinary approach to the problem of language and music specificity"), by a MINIARC grant (AT.20051024.CG.1, "Représentations mentales de la musique et du langage dans le chant et nature de leurs interactions"), as well by a FRFC grant (2.4633.66, "Mental representations of music and language in singing and nature of their interactions"). The two first authors are Senior Research Associate and Postdoctoral Researcher respectively, of FRS – FNRS. Many thanks to Thierry Vallier and Florent Hubert for their patience in singing the material used in the present studies, to Alain Soquet for his help in the stimuli analyses, to Hélène Lafontaine for testing the participants of Experiments 1 and 2, to Corine Astesano for some of the spectral analyses of the material, and to Olivier Bélanger for having carefully prepared the synthesized stimuli used in Experiment 5. We are also grateful to Emmanuel Bigand and Renée Béland for their insightful suggestions to improve this series of experiments, as well as to Steven R. Livingstone for editing the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cognition.2009.02.014.

References

Ashby, F. G., & Maddox, W. T. (1994). A response-time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38*(4), 423–466.

Astesano, C., Schön, D., & Besson, M. (2004). Le langage et la musique dans le chant. *Revue de Neuropsy News, 3*, 78–83.

Béland, R., Caplan, D., & Nespoulous, J.-L. (1990). The role of abstract phonological representations in word production: Evidence from phonemic paraphasia. *Journal of Neurolinguistics, 5*, 125–164.

Ben Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification – Role of stimulus difference. *Perception and Psychophysics, 57*, 1151–1162.

Besson, M., Faïta, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science, 9*, 494–498.

Bharucha, J. J., Curtis, M., & Paroo, K. (2006). Varieties of musical experience. *Cognition, 100*, 131–172.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction-time and musical expectancy – Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403–410.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception and Psychophysics, 59*, 1098–1107.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition, 81*, B11–B20.

Boersma, P., Weenink, D. (2007). *Praat: Doing phonetics by computer (version 5.0)*. Retrieved December 10, 2007, from <http://www.praat.org>.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science, 16*, 451–459.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens and eggs. *Psychological Science, 18*, 924–925.

Bonnel, A. M., Faïta, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception and Psychophysics, 63*, 1201–1213.

Canter, G. J., Trost, J. E., & Burns, M. S. (1985). Contrasting speech patterns in apraxia of speech and phonemic paraphasia. *Brain and Language, 24*, 204–222.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403*, 428–430.

Cohen, J., Macwhinney, B., Flatt, M., & Provost, J. (1993). Psyscope – An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods Instruments and Computers, 25*, 257–271.

Delattre, P., Liberman, A., & Cooper, F. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America, 27*, 769–773.

Escoffier, N., & Tillmann, B. (2008). The tonal function of a task-irrelevant chord modulates speed of visual processing. *Cognition, 107*, 1070–1083.

Fitts, P. M., & Biederman, L. (1965). S-R compatibility and information reduction. *Journal of Experimental Psychology, 69*, 408–412.

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

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12*, 207–222.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport, 9*, 2115–2119.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, Maryland: Erlbaum.

Garner, W. R. (1978a). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*, 98–123.

Garner, W. R. (1978b). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General, 107*, 287–308.

Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), *Perception and organization* (pp. 139–199). Hillsdale, NJ: Lawrence Erlbaum.

Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception cognition and development: Interactional analyses* (pp. 1–37). Hillsdale, NJ: Erlbaum.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225–241.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Oxford: Blackwell.

Gottwald, R. L., & Garner, W. R. (1972). Effects of focusing strategy on speeded classification with grouping, filtering, and condensation tasks. *Perception and Psychophysics, 11*, 179–182.

Gottwald, R., & Garner, W. R. (1975). Filtering and condensation task with integral and separable dimensions. *Perception and Psychophysics, 18*, 26–28.

Grau, J. W., & Kemler-Nelson, D. G. (1988). The distinction between integral and separable dimensions – Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General, 117*, 347–370.

Gregg, J., & Scherer, R. C. (2006). Vowel intelligibility in classical singing. *Journal of Voice, 20*, 198–210.

Harris, J. (2006). The phonology of being understood: Further arguments against sonority. *Lingua, 116*, 1483–1494.

Keele, S. W. (1970). Effects of input and output modes on decision time. *Journal of Experimental Psychology, 85*, 157–164.

Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage, 13*, 646–653.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*, 302–307.
Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics, 53*, 157–165.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431–461.
Lidji, P. (2007). Integrality and separability: Review and application to the interactions between lyrics and tune in songs. *L'Année Psychologique, 107*, 659–694.
Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1189–1207.
MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*, 499–546.
MacNeilage, P. F., & Davis, B. (2000). On the origin of internal structure of word forms. *Science, 288*, 527–531.
Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology – Human Perception and Performance, 22*, 795–817.
McCarthy, J. J. (1985). *Formal problems in semitic phonology and morphology*. New York: Garland.
McCrean, C. R., & Morris, R. J. (2005). Comparisons of voice onset time for trained male singers and male nonsingers during speaking and singing. *Journal of Voice, 19*, 420–430.
Mehler, J., Peña, M., Nespor, M., & Bonatti, L. L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex, 42*, 846–854.
Melara, R. D., & Marks, L. E. (1990). Dimensional interactions in language processing – Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning Memory and Cognition, 16*, 539–554.
Miller, J. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics, 24*, 175–180.
Monaghan, P., & Shillcock, R. (2003). Connectionist modelling of the separable processing of consonants and vowels. *Brain and Language, 86*, 83–98.
Morin, R. E., Forrin, B., & Archer, W. (1961). Information processing behavior: The role of irrelevant stimulus information. *Journal of Experimental Psychology, 61*, 89–96.
Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98*, 13–30.
Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical decision. *Cognitive Development, 22*, 271–279.
Nespor, M., Peña, M., & Mehler, J. (2003). On the different role of vowels and consonants in language processing and language acquisition. *Lingue e Linguaggio, 2*, 221–247.
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance: II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*, 58–117.
Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America, 119*, 1727–1739.
Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunts calls in chacuma baboons (papio cynocephalus ursinus): Implications for production processes and functions. *Journal of the Acoustical Society of America, 101*, 2951–2963.
Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 191–215). London: Erlbaum.
Patching, G., & Quinlan, P. (2002). Garner and Congruence Effects in the Speeded Classification of Bimodal Signals. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 755–775.
Peretz, I. (2006). The nature of music from a biological perspective. *Cognition, 100*, 1–32.
Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688–691.
Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination – A neuropsychological perspective. *Quarterly Journal of Experimental Psychology Section A – Human Experimental Psychology, 46*, 301–325.
Peretz, I., Kolinsky, R., Labreque, R., Tramo, M., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*, 1283–1301.
Peretz, I., Radeau, M., & Arguin, M. (2004). Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory and Cognition, 32*, 142–152.
Pinker, S. (1997). *How the mind works*. New York London: W.W. Norton and Company.
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymetric sampling in time". *Speech Communication, 41*, 245–255.
Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*, 516–540.
Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception and Psychophysics, 14*, 565–569.
Pomerantz, J., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 422–435.
Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception and Psychophysics, 18*, 355–361.
Posner, M. I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review, 71*, 491–504.
Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition, 94*, B67–B78.
Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 57*, 583–592.
Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech-perception – A cross-linguistic study. *Journal of Phonetics, 18*, 481–495.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.
Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106*, 975–983.
Schön, D., Gordon, R., & Besson, M. (2005). Musical and linguistic processing in song perception. In G. Avanzini, S. Koelsch, L. Lopez, & M. Majno (Eds.). *The neurosciences and music II: From perception to performance* (Vol. 1060, pp. 71–81). New York: Annals of the New York Academy of Sciences.
Scotto di Carlo, N. (1993). Diction et musicalité. *Médecine des Arts, 5*, 4–11.
Scotto di Carlo, N. (2007a). Effect of multifactorial constraints on opera-singing intelligibility (I). *Journal of Singing, 63*, 1–13.
Scotto di Carlo, N. (2007b). Effect of multifactorial constraints on opera-singing intelligibility (II). *Journal of Singing, 63*, 559–567.
Scotto di Carlo, N., & Germain, A. (1985). A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica, 42*, 188–197.
Selkirk, E. O. (1982). The syllable. In N. Smith & H. van der Hulst (Eds.), *The structure of phonological representations* (pp. 337–384). Dordrecht: Foris Publications.
Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Semenza, C., Bencini, G. M., Bertella, L., Mori, I., Pignatti, R., Ceriani, F., Cherrick, D., & Caldognetto, E. M. (2007). A dedicated neural mechanism for vowel selection: A case of relative vowel deficit sparing the number lexicon. *Neuropsychologia, 45*, 425–430.
Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition, 16*, 285–303.
Shepp, B. E., Barrett, S. E., & Kolbet, L. L. (1987). The development of selective attention: Holistic perception versus resource allocation. *Journal of Experimental Child Psychology, 43*, 159–180.
Sinnott, J. M., & Gilmore, C. S. (2004). Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perception and Psychophysics, 66*, 1341–1350.

Sinnott, J. M., & Williamson, T. L. (1999). Cab macaques perceive place of articulation from formant transition information? *Journal of the Acoustical Society of America, 106*, 929–937.
Sommers, M. S., Moody, D. B., Prosen, C. A., & Stebbins, W. C. (1992). Formant frequency discrimination by Japanese macaques (Macaca fuscata). *Journal of the Acoustical Society of America, 91*, 3499–3510.
Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
Sundberg, J. (1982). Perception of singing. In D. Deutsch (Ed.), *The psychology of music* (pp. 59–98). New York: Academic Press.
Thibaut, J. P., & Gelaes, S. (2002). The development of analytic and holistic processing. *L'Année Psychologique, 102*, 485–522.
Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes – Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81*, 755–764.
Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream – Functional differences between vowels and consonants. *Psychological Science, 19*, 137–144.
Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics, 15*, 501–508.
Wood, C. C. (1975). Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance, 104*, 3–20.
Zatorre, R., Belin, P., & Penhune, V. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*, 37–46.
Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society, 23*, 689–708.