Acta Psychologica 155 (2015) 43–50

Contents lists available at ScienceDirect

Acta Psychologica

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

Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception

Nia Cason [a,b,*], Corine Astésano [c,d,1], Daniele Schön [a,b,2]

[a] Aix-Marseille Université, Institut de Neurosciences des Systèmes, Marseille, France
[b] INSERM, U1106, Marseille, France
[c] UMR 7309, Laboratoire Parole et Langage, CNRS & Aix-Marseille University, 5 avenue Pasteur, 13006 Aix-en-Provence, France
[d] EA 4156, U.R.I. Octogone-Lordat, 5 allées Antonio Machado, 31058 Toulouse Cedex 09, France

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 1 December 2014
Accepted 3 December 2014
Available online xxxx

Keywords:
Speech
Music
Prosody
Metre
Rhythm

ABSTRACT

Following findings that musical rhythmic priming enhances subsequent speech perception, we investigated whether rhythmic priming for spoken sentences can enhance phonological processing – the building blocks of speech – and whether audio–motor training enhances this effect. Participants heard a metrical prime followed by a sentence (with a matching/mismatching prosodic structure), for which they performed a phoneme detection task. Behavioural (RT) data was collected from two groups: one who received audio–motor training, and one who did not. We hypothesised that 1) phonological processing would be enhanced in matching conditions, and 2) audio–motor training with the musical rhythms would enhance this effect. Indeed, providing a matching rhythmic prime context resulted in faster phoneme detection, thus revealing a cross-domain effect of musical rhythm on phonological processing. In addition, our results indicate that rhythmic audio–motor training enhances this priming effect. These results have important implications for rhythm-based speech therapies, and suggest that metrical rhythm in music and speech may rely on shared temporal processing brain resources.

1. Introduction

Though speech and music may seem distinct from one another, both are hierarchically-organised, rule-based systems whose processing shares a wide range of similarities (Jäncke, 2012; Patel, 2011). Here, the current focus is how cross-domain similarities in the processing of 'rhythm' can explain the potential for musical rhythm to impact on speech processing. Rhythm (in both domains) can be broadly defined as the temporal organisation of acoustic events which unfold over time. Two basic properties of rhythm can be said to be its ability to give rise to a sense of 'beat' – a series of regular and recurrent psychological events (Cooper & Meyer, 1960) – and 'metre', which can be described as an emergent temporal structure that results in a hierarchical organisation of salient and less salient events (London, 2012a, 2012b). The present focus is with metre: how musical metre can inform a listener about speech metre, and the consequences of this in terms of speech processing.

Western music typically has a binary, march-like metre (1**2** 1**2** 1**2**) or a ternary, waltz-like metre (12**3** 12**3** 12**3**). The emergence of metre from a sequence of rhythmic sounds is not only automatic (London, 2012a, 2012b), but its perception also said to be present from infanthood (Hannon & Johnson, 2005). This preference for perceiving metrical patterns has long been acknowledged: even on hearing an isochronous sequence of identical sounds, we automatically project a metrical structure onto them, engendering a perceptual illusion of weak and strong elements (Bolton, 1894).

Similarly in speech, salient and less salient syllables form the metrical patterning of utterances. Though speech does not possess the same degree of regularity as music (Patel, 2008), metrical 'rules' allow for a degree of rhythmic predictability in speech; a final stressed syllable marks the end of word groups in French, for example (Hirst & Di Cristo, 1998). We will refer here to stress as to the relative emphasis that is not uniquely signalled by intensity changes, but also importantly by changes in pitch and vowel duration. There are multiple levels of saliencies in speech rhythm: there are different *types* of prosodic stress (e.g. lexical stress, pitch stress, and emphatic stress), all of which have different functions in speech, as well as different *degrees* of stress (e.g. primary and secondary lexical stress; and high, low or complex pitch stress). As in the music domain, it is the interactions between these multiple levels of stress that allow for the emergence of speech rhythm (Handel, 1989: pp. 383).

Abbreviations: A group, auditory modality exposure group; AM group, audio–motor training group; RTs, reaction times.

* Corresponding author at: INSERM, U1106, Marseille, France. Tel.: +33 4 91 38 55 77.
E-mail addresses: niacason@yahoo.co.uk (N. Cason), corine.astesano@univ-tlse2.fr, corine.astesano@lpl-aix.fr (C. Astésano), daniele.schon@univ-amu.fr (D. Schön).
[1] Tel.: +33 5 61 50 24 69.
[2] Tel.: +33 4 91 38 55 77.

http://dx.doi.org/10.1016/j.actpsy.2014.12.002
0001-6918/

In both domains, metre is key in allowing a listener to form predictions about what will happen next. For instance, on hearing a ternary metrical sequence in music (12**3** 12**3** 12**3**), we are able to automatically and implicitly predict that a '12**3**' (weak–weak–strong) pattern will follow. This is also the case in speech, whereby hearing a list of trisyllabic words with a final stress induces expectations for a word with the same metre (Pitt & Samuel, 1990). Given this, it is perhaps not surprising that attentional accounts have been used to explain predictive mechanisms in both speech (Pitt & Samuel, 1990) and in music (Jones, 2008; Large & Palmer, 2002). In speech, The Attentional Bounce hypothesis states that attention is oriented to syllables which are expected to be stressed (Pitt & Samuel, 1990). It claims that the position of these stressed syllables can be predicted on the bases of the metrical patterns of speech, and that this is reflected by quicker phoneme detection at attended syllables (Pitt & Samuel, 1990; Rothermich & Kotz, 2011). In music, the Dynamic Attending Theory (DAT) predicts that, on hearing a metrical rhythm (external oscillators), neural rhythms synchronise to it with similar phase and period relations (internal oscillators), and that this coupling dynamically modifies attention in time, with more attentional resources being allocated to strong (predictable) metrical positions (Jones, 2008; Large & Jones, 1999; London, 2012a, 2012b; Snyder & Large, 2005). As a result, auditory events occurring at these more predictable locations are better processed (Barnes, Barnes, & Jones, 2000; Ellis & Jones, 2010; Jones, Boltz, & Kidd, 1982; Jones, Moynihan, MacKenzie, & Puente, 2002). This effect of auditory rhythmic expectation is also cross-modal, and can enhance the detection of visual targets, too (Brochard, Tassin, & Zagar, 2013; Escoffier, Sheng, & Schirmer, 2010).

To conclude thus far, metre can be said to dynamically modulate attention in time in such a way that processing efficiency is influenced. This possibly takes place is a similar way in music and speech, as shown by impacts of musical metre on the phonological processing of speech (Cason & Schön, 2012). Should this be the case, expectations induced by musical metre can be hypothesised to impact on the processing of speech metre, and, as a consequence, to impact on the phonological processing of speech. Notably, this prediction is scale free: it considers only metrical relations, which are considered to include 'number' expectations (how many elements to expect) and 'stress' expectations (what stress patterning to expect), and not the exact durational expectations (i.e. interonset intervals).

This hypothesis (that musical rhythm can have a cross-domain impact on speech processing) has been received convincing support. For instance, temporal alignment between stressed beats and stressed syllables allows for a greater comprehension of speech (Gordon, Magne, & Large, 2011). More specifically in relation to the current research question, inducing rhythmic expectations can have also have a cross-domain effect on *subsequent* speech processing (Cason & Schön, 2012). In that experiment, expectations about beat and metre were induced by a musical rhythmic prime, which was followed by a bi- or trisyllabic pseudoword with a final stressed syllable. Focussing here on the metre aspect, we found electrophysiological (EEG) evidence of an enhanced processing when the speech metre matched the prime metre (for example, when a prime with a ternary metre (12**3** 12**3** 12**3**) was followed by a trisyllabic word with the same ternary metre (12**3**)). These results indicate that musical rhythm impacts on phonological processing (stimuli for this experiment were pseudowords with no semantic meaning) by inducing domain-general metrical expectations. To build upon these findings, the current study aimed to investigate whether this effect – seen at the level of single pseudowords – would also be present for real sentences.

Because active (audio–motor) engagement with musical rhythm may allow listeners to further internalise musical rhythms which can be shared by speech, a second aim of the present study was to investigate the role of rhythmic sensorimotor training on this cross-domain priming effect. In both speech and music, audio–motor interactions are vitally important for the consolidation of accurate production and the continual monitoring of output (Guenther, 2006; Lappe, Herholz, Trainor, & Pantev, 2008; Levelt, 1999; Zatorre, Chen, & Penhune, 2007). In music, audio–motor training can strongly influence the perception of rhythm (Phillips-Silver & Trainor, 2005, 2007; Su & Pöppel, 2012) and can enhance listeners' sensitivity to metrical deviants (Geiser, Sandmann, Jäncke, & Meyer, 2010; Vuust et al., 2005). Considering this, rhythmic training would intuitively result in a stronger priming effect, through an enhanced metrical sensitivity. From a clinical standpoint, the importance of musical audio–motor training in enhancing speech fluency has not been fully acknowledged. In therapies such as Rhythmic Speech Cueing (Thaut, 2005), for example, it is often the therapist who produces the rhythm on a drum whilst the patient is asked only to produce speech (rather than themselves actively participating in the musical rhythm production). In this case, any further benefits of motor rhythmic engagement are not fully potentiated.

In the present experiment, a musical metrical prime sequence was used to induce metrical expectations about both stress patterns and the number of elements. A subsequently-heard sentence either conformed to these stress and number expectations – and thus to listeners' metrical expectations – or did not. Considering the impact of rhythmic predictability on processing efficiency, we hypothesised that in conditions where metrical expectations were met, phonological processing of a target sentence would be enhanced. The experiment employed a phoneme detection task to measure phonological processing, a building block of speech perception, with faster reaction times (RTs) indicating a facilitated access to phonological information. To address the second aim of this study, we tested two groups: one group who underwent audio–motor training with the musical rhythms presented in the experiment (Audio–Motor (AM) Group), and another group without this training (Auditory only (A) Group). We hypothesised that the positive effect of metrical priming on phonological processing would be more striking in the AM Group due to the consolidation of metrical representations afforded by audio–motor training.

2. Materials and methods

2.1. Stimuli

Experimental trials consisted of a rhythmic prime sequence followed by a sentence. We used four rhythmic prime sequences that comprised four experimental blocks (Prime 1, Prime 2, Prime 3, Prime 4), and four sentence types built around these same rhythms (Fig. 1).

Prime sequences were created in Adobe Audition and differed in number of beats (6 or 7) and in the placement of two stressed events. These rhythmic patterns were composed of percussion sounds, which had a stimulus onset asynchrony of 225 ms (beat level = 450 ms). The first stressed percussion sound of each prime had a rim shot timbre, a duration of 196 ms and an average Root Mean Square (RMS) of −41.25 dB, the second stressed percussion sound of each prime had a snare timbre, a duration of 353 ms and an RMS of −28.88 dB, and unstressed percussion sounds had a closed high hat timbre, a duration of 138 ms and an RMS of −43.63 dB.

Sentence stimuli were constructed around the four prime rhythms, thus using two syllable groups of 3 or 4 syllables (the most common metre in French, whatever the speaking style, Astésano, 2001). 40 sentences were constructed for each of the four prime rhythms, resulting in a total of 160 sentences. The final word of the sentence (in which the phoneme target was present 50% of the time) was selected using Lexique 3 (New, Pallier, Ferrand, & Matos, 2001). These sentence-final words were bisyllabic, had a CV/CV structure (at the phonological level), and were balanced for lexical frequency across the four sentence rhythm types. The last syllable of each word (an open CV syllable) contained a target vowel, for instance, as/i/in/*mari*/ for 'mari', and the previous consonant macro-class (liquid, nasal, unvoiced fricative, voiced fricative, unvoiced stop and voiced stop) was balanced over the 4 rhythmic conditions.