Temporal Coordination and Prosodic Structure in Autism Spectrum Disorder: Timing Across Speech and Non-Speech Motor Domains

Abstract

Individuals with autism spectrum disorder (ASD) often exhibit disordered speech prosody, but sources of disordered prosody remain poorly understood. We explored patterns of temporal alignment and prosodic grouping in a speech-based metronome repetition task as well as manual coordination in a drum tapping task among Cantonese speakers with ASD and normal nonverbal IQ and matched controls. Results indicate similar group results for prosodic grouping patterns, but significant differences in relative timing and longer syllable durations at phrase ends for the ASD group. Variability on the speech task was significantly correlated with variability on the drumming task, consistent with the view that impairment in both speech and non-speech motor domains can be linked with deficits in temporal processing.

Temporal Coordination and Prosodic Structure in Autism Spectrum Disorder: Timing Across Speech and Non-Speech Motor Domains

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in communication and social interaction (American Psychiatric Association, 2013), as well as by impairments in motor control (Bhat et al., 2011). In terms of communicative deficits, disordered prosody— which refers primarily to suprasegmental features of speech such as phrasing, intonation, and rhythm—have been noted since early work on ASD (Asperger & Frith, 1991; Kanner, 1968). Clinical impressions have variously referred to the speech of individuals with ASD as 'monotone', 'robotic', 'staccato', 'jerky', and 'sing-songy' (Asperger, 1952; Baron-Cohen & Staunton, 1994; Kanner, 1943), and atypical prosodic features have been described as among the first identifiable characteristics that create an impression of "oddness" among peers of individuals with ASD (Mesibov, 1992; Van Bourgonien & Woods, 1992). A growing body of acoustic studies have confirmed such impressions, demonstrating that individuals with ASD show generally slower speech rate (Patel et al., 2020), greater intonational range (Diehl & Paul 2013; Nadig & Shaw, 2012) as well as differences in durational cues to prosodic phrasing and stress (Fosnot & Jun, 1999; Paul et al., 2008). The two gold-standard diagnostic tools of ASD, Autism Diagnostic Observation Schedule, second version (ADOS-2) (Lord et al., 2012) and Autism Diagnostic Interview-Revised (ADI-R) (Le Couteur et al., 2003) both include speech prosody impairments as a diagnostic feature of ASD.

There is still considerable work to be done in understanding the nature of disordered prosody in ASD. In particular, few studies have attempted to understand speech prosodic deficits in the context of more general deficits in motor control in the disorder. Evidence has been increasing which suggests that motor deficits linked to temporal processing affect a large number of people on the autism spectrum, including those who are very high functioning (Kaur et al., 2018). Therapeutic interventions which focus on rhythm and timing stability have been found to improve outcomes in ASD across both the speech and motor domains (Berger, 2012; Sharda et al., 2018; Srinivasan et al., 2015). Despite this, there has been little work that attempts to directly link speech-related deficits in ASD with other motor deficits, especially within the domain of speech prosody. Here, we attempt to close some of that gap by exploring parallels between speech timing and timing on a drumming task among Cantonese-speaking individuals with ASD. Specifically, we compare performance on a metronome-guided *speech cycling task*--in which participants repeat sentences in time to a metronome set to different rates—with those of a similarly rate-controlled drumming task. This provides us with a parallel task structures across speech and non-speech domains, and also allows us to observe coordinative behavior within a tightly controlled context. Furthermore, the structure of the speech cycling task allows us to evaluate a number of different aspects of prosodic timing—from relative alignment of prosodic phrase boundaries to variation in syllable duration—which will help to provide a more detailed account of which aspects of prosody are most impacted in ASD. In the sections that follow, we provide an overview of the various cross-linguistic functions of speech prosody, and summarize key findings on deficits in timing and motor control in ASD, before turning to our study methods and results.

Cross-linguistic Functions of Prosody and the Impact of ASD

Understanding disordered prosody in ASD is complicated by the fact that prosody is implicated in a number of different aspects of linguistic function, and prosodic marking can vary considerably by language. For example, pragmatic information concerning which elements of discourse are under discussion can be encoded prosodically. In English, pitch and duration differences can be used to indicate which word in a sentence is contrastively focused, allowing for a distinction to be made between the sentence Mary saw SUE (where 'Sue' has greater pitch and duration and is interpreted to contrast with someone else whom Mary might have seen) and MARY saw Sue (where 'Mary' has greater pitch and duration and is interpreted to contrast with someone else who might have seen Sue). In tone languages, focus can also be prosodically-marked, though in ways distinct from non-tonal languages like English. In Cantonese, for example, focus is marked by means of a sentence-final particle whose pitch patterns and duration convey additional information about sentence type (e.g. question vs. statement), speaker affect, and epistemic context (Li, 2006). In either language, conversation partners must infer certain things about the discourse situation and share a certain amount of common ground information in order for these aspects of prosody to be successfully interpreted. A number of studies have found that individuals with ASD have trouble producing and interpreting these types of pragmatically-based prosodic cues, and many researchers have interpreted this as further evidence of a social-communicative deficit in ASD (Chan & To, 2016; Diehl & Berkovits, 2010; McCann & Peppé, 2003).

Not all aspects of prosody are directly linked with social-pragmatic aspects of language use, however. Prosody can also serve a different function which is organizational in nature. For example, English organizes syllables into maximally disyllabic *feet* in which one syllable (the 'head') carries stress, meaning it has longer duration and higher amplitude than the other (Fry, 1955; Liberman & Prince, 1977). The organizational role of prosody can also be observed for larger phrasal constituents, as well: in both English and Cantonese (as well as in many other languages), syllables immediately preceding a phrase boundary are lengthened (or ‘slowed’) as compared with syllables which occur phrase-internally, and pauses are longer before the start of a new phrase than in the middle of a phrase (Beckman & Edwards, 1990; Wong et al., 2005). While these types of temporal variations may be associated in some cases with grammatical distinctions (such as the pauses which differentiate the restrictive versus non-restrictive English relative clauses *the child who is in the hall…* and *the child, who is in the hall,…*), there is not always a straightforward mapping between prosodic timing and grammatical function (Nespor & Vogel, 1986; Selkirk, 1984). It has therefore been hypothesized that such aspects of prosodic organization and grouping are more closely tied with a functional demand to break speech up into smaller ‘chunks’ for the purposes of speech planning and auditory processing (Ferreira, 1993; Shattuck-Hufnagel & Turk, 1996).

There is some research which suggests that these non pragmatically-based, organizational aspects of prosody are also impacted in ASD. For example, Paul et al. (2008) showed that English-speaking individuals with ASD showed less durational variability across stressed and unstressed syllables in shadowed nonsense words. In a sentence completion task, Grossman et al., (2010) found that individuals with ASD produced lexical items in phrase-final position with atypically long durations. Fosnot & Jun (1999) also found that individuals with ASD produced longer durations than neurotypical controls when producing identical phrases. A recent ERP study by DePriest et al. (2017) found that individuals with ASD show less sensitivity to pauses at phrase boundaries. Taken together, these results suggest that prosody is impacted not only within the socio-pragmatic functional domain, but in the organizational domain, as well.

Impaired Temporal Processing as a Source of Prosodic Impairments in ASD?

While differences in pragmatically-based prosodic functions such as contrastive focus may be attributable in ASD to difficulties with social-cognitive abilities including 'theory of mind' (Baron-Cohen et al., 1985), such an explanation cannot adequately capture differences in prosody which are not straightforwardly linked to pragmatic function (e.g. the decreased durational variability on nonsense words documented by Paul et al., 2008). Where, then, do such differences stem from? One possibility is that these patterns—many of which crucially involve deviations in speech timing—originate in a more general temporal processing deficit in ASD. Indeed, there is mounting evidence to suggest that impaired representation of timing is a common deficit in ASD. For example, using a temporal bisection task, in which subjects compare the duration of a stimulus with one held in memory, Allman et al. (2011) found that individuals with ASD had points of subjective temporal equality which were significantly different than those of their neurotypical (NT) peers. Using a similar task, studies by Martin et al. (2010) and Szelag et al. (2004) found that individuals with ASD reproduced tone durations less accurately than NT controls, usually reproducing tones with durations which were too short. Szelag et al. also included a condition with a visual stimulus, rather than an auditory one, and found similar results. These findings suggest that temporal processing deficits in ASD do not simply reflect poor auditory processing.

Given the fundamental role that timing plays in so many aspects of human functioning—including many areas which are impaired in autism—some have argued that poor timing ability, rather than being peripherally associated with core deficits in autism, in fact plays a more direct, causal role in determining behavioral patterns in the disorder (Allman, 2011; Allman & deLeon, 2009; Allman & Falter, 2015). In the domain of language, there is evidence that deficits in temporal estimation may be directly implicated in speech perception difficulties in ASD. For example, EEG research by Lepistö et al. (2005, 2006) revealed abnormal patterns in mismatch negativity (MMN)—an event-related potential elicited by an odd stimulus presented in a sequence of otherwise identical stimuli—in response to durational variability in speech sounds. The unique role of temporal processing deficits was highlighted in the same study by the fact that participants with and without ASD performed comparably in response to pitch variations on a similar task. Further work by Alcantara et al. (2004) revealed that individuals with autism had more difficulty perceiving speech in noise than NT controls, but that these difficulties were especially prevalent for noise accompanied by 'temporal dips,', or momentary fluctuations in the signal-to-noise ratio; such fluctuations are common at prosodic boundaries in speech production. A study by Groen et al. (2009) had similar findings, and showed that noise involving temporal modulations was more detrimental to processing in ASD than noise involving spectral modulations.

The importance of temporal processing is of course not limited to the perceptual domain. Current models of motor control—including for speech—posit that feedforward mechanisms for joint and muscle control rely on predictive processes which allow an individual to rapidly estimate the state of one joint or muscle in order to accurately coordinate timing of other joints and muscles involved in the same motor task (Todorov & Jordan, 2002; Tourville & Guenther, 2011; Parrell et al., 2019). Temporal estimation for feedforward programs has been linked to two subcortical regions—the cerebellum and the basal ganglia (Albin et al., 1995; Parrell et al., 2017; Pickett et al., 1998; Van Buren, 1963)—both of which show irregularities in individuals with ASD. Research examining the anatomy of the cerebellum has found a high incidence of abnormality in the cerebellar vermis and hemispheres in individuals with ASD as compared with neurotypical controls (Rogers et al., 2013). It has been hypothesized that these types of cerebellar abnormalities may help to explain observed patterns of decreased timing accuracy in ASD (Gowen & Miall, 2005) and may contribute to increased rates of 'clumsiness' and gait disruption in the disorder (Hallett et al., 1993; Vernazza-Martin et al., 2005; Rinehart et al., 2006). Studies of the structures within the basal ganglia have revealed reduced white matter volume and decreased functional anisotropy of white matter tracts between the putamen and frontal cortex among adults with ASD (Langen et al., 2012), as well as decreased volume of the globus pallidus (Sussman et al., 2015). These abnormalities in the basal ganglia have been directly linked with repetitive and stereotyped motor behaviors among individuals with ASD (Estes et al., 2011).

Despite clear parallels in disruptions in speech and other motor behavior in ASD, particularly where timing of movement and speech is concerned, few studies have investigated disordered speech prosody from the broader context of impaired motor control. As such, many studies have incorporated fairly coarse-grained acoustic measures of whole word or even whole phrase durations, without attention to syllable-by-syllable timing within a phrase. As a result, these studies cannot highlight which aspects of speech timing are most heavily impacted. For example, longer phrase durations observed in ASD (e.g. by Fosnot and Jun, 1999) could reflect an overall slower speech rate, or could reflect more dramatic application of common prosodic processes such as phrase-final lengthening, where the syllables toward the end of a sentence are lengthened in duration relative to earlier syllables (Byrd & Saltzman, 1998; Pierrehumbert, 1980). Without looking at relative timing and duration of syllables across positions within a phrase, it is hard to disentangle these possibilities. Determining the true patterns is not only important from a descriptive perspective, but also has implications for our understanding of underlying impairments in ASD: while overall slower speech rate may be consistent with an impairment in working memory abilities (Hulme et al., 1984; Baddeley, 1986, 1990), more localized disruptions in speech timing may point to problems with motor planning or implementation. A further possibility is that grammatical deficits may be at play: for example, research has found that individuals with ASD struggle to use prosodic phrase boundaries as cues to interpret syntactically ambiguous sentences (Diehl et al., 2008); it is possible that phrasing is similarly impaired in speech production. We explore all of these possibilities in the current study.

Examining Timing Across Motor Domains: Speech Cycling and Drumming Tasks

Speech Cycing task

The speech cycling task, originally developed by Cummins (1997) and Cummins & Port (1998), was designed to elicit highly regular speech samples with the primary goal of evaluating how speakers of different languages organize speech prosodically in time while controlling for speech rate. In the task, participants repeat short phrases in time with a metronome set to progressively shorter periods/faster speech rates. Phrases are repeated once per metronome beat (Figure 1).

[FIGURE 1 ABOUT HERE]

Of particular interest is the *relative timing* within the repetition cycle at which syllables in the phrase are aligned, particularly as the prosodic structure of the phrase is manipulated. Results from a variety of languages using the task have revealed that, when given short (4-6 syllable) sentences, speakers will group the phrase into smaller prosodic units. Across phrases of identical lengths (in syllables), syllables occurring in initial position of a prosodic unit tend to be timed later than those occurring in the same numerical position which do not initiate a new unit (Figure

2) (Tajima, 1998; Franich, 2017, 2019). This fact reflects the tendency across languages to pause at prosodic boundaries and lengthen the final syllables of prosodic phrases (particularly where they align with a syntactic phrase boundary), as well as rhythmic constraints which determine that ‘heads’ of prosodic units (the rhythmically strongest syllable in the unit which drives the sense of a ‘beat’ or ‘pulse’ during speech) are drawn to specific, lower-order proportions of the repetition cycle, such as the halfway (.5) point.

[FIGURE 2 ABOUT HERE]

Pause length between prosodic units has been shown to vary as a function of preceding and following phrase length; generally, the longer the prosodic phrases preceding and following the pause, the longer the pause will be (Grosjean et al., 1979; Krivokapić, 2007). Thus, since structures like the 2-2-2 pattern in Figure 2 have more prosodic phrases but fewer syllables per phrase, pauses are predicted to be more frequent but shorter within the phrase than those found for the corresponding 3-3 structure. In all three structures, syllables at the start of a new phrase (indicated with brackets around syllables) are predicted to occur later than corresponding syllables which do not initiate a phrase, and closer to ‘harmonic’ phase positions, such as the halfway point (indicated with dotted gray lines).

In languages like English which feature a disyllabic foot unit which conditions the position of lexical and phrasal stress, the foot plays an integral role in determining relative timing in the speech cycling task (Cummins & Port, 1998; Tajima, 1998). In Cantonese, a lexical tone language with no clear evidence of stress or foot structure, it has been shown that temporal alignment strategies in the speech cycling task are driven by other aspects of prosodic phrasing, such as the position of *intermediate phrase* (IP) boundaries (Yip, 2018). Intermediate phrase boundaries typically align with the edges of syntactic phrases such as a noun phrase or verb phrase (Pierrehumbert and Beckman, 1986) (Figure 3). Specifically, it has been shown that speakers align syllables initiating intermediate phrases at later positions than syllables occurring in phrase-medial position, even when they occur in the same numerical position (e.g. σ_1, σ_2, σ_3, etc.) within the sentence. These results are in agreement with work by Chow et al. (2010) which posited that phrase boundaries, rather than foot or word boundaries, were the most relevant for determining Cantonese prosodic organization.

In addition to examining inter-phrase prosodic organization through relative timing in the speech cycling task, we also aim to explore another aspect of prosodic timing, that of phrase-final lengthening. Specifically, speakers of a wide variety of languages are found to display considerable lengthening of syllables occurring at the ends of *intonational phrases*, prosodic units which tend to align with clause- or sentence-level syntactic constituents (Pierrehumbert, 1980) (Figure 3). Intonational phrase-final lengthening is notable in that it tends to be even more pronounced than lengthening at smaller phrase boundaries, such as intermediate phrases (Byrd & Saltzman, 1998). In some languages, including several Chinese languages, intonational phrase-final lengthening has been found to affect not just the final syllable, but to apply to penultimate and antepenultimate syllables as well, with syllables increasing continuously in duration as they approach the intonational phrase boundary (Lin & Fon, 2010; Fon et al., 2011).

[FIGURE 3 ABOUT HERE]

The benefit of using the speech cycling paradigm in the present study is that it allows us to observe speech in a highly structured setting in which factors such as speech rate are explicitly controlled for. The task also minimizes interpersonal demand, allowing participants to focus on the task of speaking.

Drumming task

The speech cycling task provides us with a speech-based coordination task which lends itself to comparison with tasks involving other types of motor coordination, such as manual drumming. In order to assess the relationship between coordination, prosodic patterns, and coordinative timing more generally, we had participants perform a separate task in which they tapped on a drum pad with a stick at the same metronome speeds used for the speech cycling task. Drumming and tapping—two types of motor entrainment tasks—are common tasks used to evaluate motor control, including among patients with Parkinson's disease and other motor-related disorders (Jobbágy et al., 2005). Previous work has already demonstrated that individuals with ASD show impairments in motor entrainment (Gowen & Miall, 2005; Morimoto et al., 2018), a fact which has been attributed to impaired timing due to cerebellar disfunction in ASD. As pointed out by Morimoto et al., a benefit of this type of task in the context of autism research is that performance on the task doesn't covary with IQ to the degree that other motor tasks do (Martin et al., 2010; van der Fels et al., 2015); it is therefore both a sensitive and specific measure of ASD-related motor difficulties. At the same time, performance on motor entrainment tasks has been explicitly linked with other motor function impairments in ASD (Gowen & Miall, 2005), suggesting that this task provides a reasonable general measure of motor skills in ASD.

Study predictions

We predict that participants will generally demonstrate evidence of prosodic grouping in the speech cycling task, such that syllables initiating a new intermediate phrase will be timed slightly later than those in the same numerical position which do not initiate a new phrase. We also predict, in line with previous findings, that participants with ASD will exhibit less accuracy in metronome alignment in the drumming task. Assuming that a more general timing deficit exists for those with ASD, we predict that those participants will also display greater difficulty aligning with the metronome in the speech cycling task. An additional question concerns whether ASD participants will show differences from controls in prosodic alignment patterns in the speech cycling task. Note that, since the measure of temporal alignment used in the speech cycling is based on a calculation of relative timing within the speech repetition cycle, it is possible for speakers to show poor alignment with the metronome while still showing typical patterns of prosodic alignment among syllables within the phrase. In the event that accuracy in metronome alignment is governed by the same processes which guide prosodic phrasing, we expect that alignment accuracy and prosodic phrasing will be similarly impacted among participants with ASD. Should the two types of timing rely on different mechanisms, we might expect that individuals with ASD will excel at prosodic grouping while having difficulties with metronome coordination. A similar question holds for intonational phrase-final lengthening: if this type of lengthening is governed by similar timing mechanisms as for metronome coordination, we expect that performance on metronome coordination will predict the degree of intonational phrase-final lengthening found for participants.

Method

Participants

Ten native Cantonese speakers with ASD between the age of 22 and 31, and ten NT native Cantonese speakers participated in the study. All the participants with ASD have received a formal diagnosis of ASD from either a clinical psychologist or a pediatrician before adulthood. The current status of ASD was verified by the clinical judgment of the last author who is a speech-language pathologist with ASD expertise, and the Autism Diagnostic Observation Schedule (Lord et al., 2012) administered by research-reliable personnel, with a total score at or above the thresholds of autism or autism spectrum for Module 4. The nonverbal IQ of the participants with ASD was assessed with the Test of Nonverbal Intelligence (TONI-4; Brown et al., 2010). They displayed average to above average level with the mean index score of 102.33 (SD = 9.56). The hearing ability of all participants was screened with a GSI 18 screening audiometer in a sound-proofed room, with the passing criteria set at 25 dB HL at the frequencies of 1000, 2000 and 4000 Hz in both ears (ASHA, 2020). Due to the possible influence of musical training on the participants' timing abilities (Miendlarzewska & Trost, 2014), besides age, sex, and education level, musical experience was included as a matching parameter for the two groups. There were 9 males and 1 female in each of the two groups, and they were all receiving or have attained post-secondary level education at the time of experiment. Independent-sampled t-test indicated that there were non-significant differences between the mean of their age, $t(18) = 0.35$, $p = .74$, and years of musical training received, $t(18) = 0.08$, $p = .94$, in both groups.

Procedures

There were two experimental tasks: (1) speech cycling task and (2) tapping task. The procedures were the same for the two groups of participants. The speech cycling task adopted the paradigm in Yip (2018), which was originated from the speech cycling task developed by Cummins (1997). The paced drum beating task developed in Corriveau and Goswami (2009) was referenced for the tapping task.

All the recordings took place at a soundproof booth at The University of Hong Kong. They spoke into a microphone (Shure SM 58). They wore headphones (Sennheiser HD 280 pro) through which the metronome beats were played. The metronome sound consisted of a synthetic drumbeat created in version 2.1.2 Audacity® recording and editing software, an open-source program for sound editing. At the beginning of the speech cycling task, written instructions along with auditory instructions were presented via PowerPoint slides. The study materials developed by Yip (2018) were used. The items were shown on the screen one by one in random order. The materials consisted of 15 sentences with five different prosodic structures, with intermediate phrase boundaries predicted to occur at the edges of syntactic phrases. All the syllables in the stimuli are high-level tone in order to control for any possible influence of pitch on the speech rhythm. Table 1 shows a sample of each syntactic structure from the materials used.

On every trial, four beats of metronomes were first presented to allow speakers to familiarize with the rate. They were then asked to read aloud the sentence in time with the metronome at least five times without taking a breath. No form of tapping was allowed during the speech cycling task. The production of each sentence started from the slowest rate (1384 ms between beats) and sped up by decreasing the duration between beats by 13.5% for each successive speed until the duration reached 774 ms. Thus, the same phrase item was repeated for five different rates in total before the participant moved onto the next phrase item. The purpose of acceleration is to examine the stability of syllable alignment pattern across speech rates. The same procedure was applied to every sentence stimulus. Before the experiment, there were two practice trials to ensure that the participants understand the procedures. Speech and drumbeats were recorded to separate audio channels so as to facilitate acoustic analysis.

[TABLE 1 ABOUT HERE]

The speech cycling task was followed by the tapping task. Participants were asked to hit a 10” electronic drum pad using a drum stick with their dominant hand along beats of the same speed as the speech cycling task. Their responses were recorded simultaneously with the beats as a stereo soundtrack. Same as the speech cycling task, four lead-in beats were presented first to alert them the tempo they would entrain with. They were instructed to hit the drum pad in time with the rhythm of the beats. Similarly, the beats were presented from the slowest tempo to the fastest tempo, then back to the slowest tempo again. Each metronome beat pace was presented for 30 seconds, followed by five seconds pause before the next beat started.

Analysis and Measures

Speech cycling task

Audio recordings were first processed with PRAAT (Boersma, 2002) to annotate the onset and termination time of each syllable and the onset timing of the beat. A trial was defined as the repetition of each six-syllable sentence under each of the specific rate. Timing information was only obtained from the first five repetitions in each trial. Given that there were a total of 15 sentences, produced at five different tempi for each sentence, with five repetitions at each rate, data on 2250 syllables (15 sentences x 5 tempi x 5 repetitions x 6 syllables) were analyzed for each participant.

Relative timing was calculated according to the notion of *external phase* from Cummins (1997) and is presented in Figure 4. For each repetition cycle, the duration of time between successive utterances of the first syllable (identified as the onset of acoustic energy in a spectrogram) was calculated (interval *a*). Then the duration from the start of the phrase until the start of the target syllable (identified based on specific spectral characteristics of the syllable, such as onset of high frequency random noise for fricative-initial syllables, onset of silent closure for stop- or affricate-initial syllables, or onset of nasal closure for nasal-initial syllables) was calculated (interval *b*). In order to derive a measure of relative timing ('proportion phase') within the repetition cycle for Syllable 6 (σ6), interval *b* would be divided by interval *a*. While the researcher annotating the data (the second author) was not blind to the diagnostic status of the participants, since the drum beats were annotated separately from the speech and tapping data, and since speech and tapping measures were calculated in relative time, it is unlikely that the researcher's annotations would have been unduly affected by her knowledge of the participants.

[FIGURE 4 ABOUT HERE]

To investigate timing accuracy and variability across groups, we examined effects of diagnostic group on two variables: *speech asynchrony* and *inter-repetition interval*. The speech asynchrony represents the absolute value of the timing difference between the production of first syllable in each repetition and the given beat immediately before that. The inter-repetition interval refers to the time interval between two consecutive repetitions of the first word in the target phrase.

Two separate linear mixed effects models were built to investigate the effect of diagnostic status on each of these variables. Each model included fixed effects of GROUP (coded as a factor two levels: ASD vs. NT) and METRONOME SPEED (coded as a numerical variable). The factor GROUP was sum coded, such that the two levels were compared with the grand mean as the intercept. Models also included by-subject random slopes for METRONOME SPEED. Following best practices outlined in Barr (2013), maximal models including random slopes for all variables were fit first; if convergence was not achieved, slope parameters were removed until convergence was reached. We calculated p-values for linear mixed effects models using Satterthwaite's degrees of freedom method, implemented in the *lmerTest* package for *R* (Kuznetsova et al., 2017). Nakagawa's R^2 is reported as a measure of variance explained by each factor, akin to a standardized effect size estimate for linear mixed effects models (Nakagawa & Schielzeth, 2012). The first number in each R^2 value reflects the conditional R^2, or the proportion of variance explained by both fixed and random factors. The second number, the marginal R^2, reflects variance explained by each fixed effect alone.

Tapping task

Audio recordings were first processed with MATLAB (The MathWorks, Inc., 2012) to extract the timing information of the drum hits by the participants and the beat they were asked to entrain to. Data on any drum hit before the fifth given beat or during the five second pause was discarded. Similar to the speech cycling task, a *tap asynchrony* was computed, reflecting the timing difference between the participant's drum hit and the given beat. We also measured the *inter-tap interval*, which refers to the time interval between two consecutive drum hits.

Two additional linear mixed effects models were built to investigate the effect of diagnostic status on each of the two tapping variables. Each model included fixed effects of GROUP (coded as a factor two levels: ASD vs. NT) and METRONOME SPEED (coded as a numerical variable). The factor GROUP was sum-coded, such that the two levels were compared with the grand mean as the intercept. Once again, models also included by-subject random slopes for METRONOME SPEED.

Relative timing and duration as measures of prosodic phrasing

Of primary interest in the study was the influence of diagnostic status on prosodic phrasing. To investigate this relationship, additional linear mixed effects models were built to evaluate variation in relative timing of syllable alignment (see Figure 4) across prosodic forms in the Speech Cycling task as a function of group. Separate models were built for each syllable for Syllables 3-5 in the speech cycling task, since these syllables were all predicted to exhibit differences in alignment based on prosodic form. For each of these models, fixed effect included factors PROSODIC FORM (5 levels, corresponding to each of the forms shown in Table 1), and GROUP (two levels: ASD vs. NT), as well as a continuous variable for METRONOME SPEED. The variable PROSODIC FORM was treatment coded with the reference level set to the 3-3 syllable grouping condition (Form #2 in Table 1). The 3-3 grouping condition was chosen as the reference level since it was predicted that the greatest differences in prosodic timing would be found between that condition and the other four conditions . Two-way interactions were included between PROSODIC FORM and GROUP. Random intercepts for Subject and Item were also included in the models, as was a by-subject random slope for METRONOME SPEED.

One additional model was created to evaluate the effects of diagnosis on intonational phrase-final lengthening. A single model was run in which syllable duration was included as the dependent variable. This model included all of the same variables as the model detailed for relative timing, but swapping out the prosodic form factor for a new factor, SYLLABLE ORDER, which included four levels for each of the four final syllables in each phrase. This factor was treatment coded, with the reference level set to Syllable 3. The reason for looking at syllables beyond just the final syllable in the intonational phrase was that, while intonational phrase-final lengthening is typically greatest for the final syllable of the phrase, it has also been found to extend to penultimate and even the antepenultimate syllable in some languages (Shattuck-Hufnagel & Turk, 1998). A two-way interaction was included in the model between GROUP and SYLLABLE ORDER. A by-subject random slope was also included for METRONOME SPEED.

Sensitivity and specificity of timing measures

In order to assess sensitivity and specific of each of our timing measures, additional logistic regression models were built with each of our six timing variables—SPEECH ASYNCHRONY, INTER-REPETITION INTERVAL, TAP ASYNCHRONY, INTER-TAP INTERVAL, RELATIVE TIMING, and SYLLABLE DURATION—as predictors of diagnostic status, with METRONOME SPEED included as a covariate.

Results

Effects of Group on timing in speech and tapping

For the first measure of speech coordination, *speech asynchrony*, a significant effect of GROUP was found (t = -5.21, β =-.11, $p < .001$, $R^2 = .09, .05$), indicating the ASD group timed their repetitions significantly further from the metronome beat than did the NT control group (Figure 5). While about half of the overall speech asynchrony scores indicated anticipation of the metronome beat as opposed to alignment or lag, ASD participants were more likely than controls to anticipate than to lag, accounting around 60% of the anticipated scores. There was also a significant effect of METRONOME SPEED ($t = 2.49$, $\beta = .03$ $p < .05$, $R^2 = .10, .004$), reflecting that participants overall showed less accurate timing with respect to the beat as speed of the metronome increased. Standard deviations for speech asynchrony were comparable across groups, around 95 ms. For the second measure of speech coordination, *inter-repetition interval*, a more modest effect of GROUP was found ($t = 2.16$, $\beta = .03$ $p < .05$, $R^2 = .11, .02$), indicating that the ASD group showed slightly longer intervals between utterances of the first syllable of the phrase (Figure 6). The ASD group also showed a considerably higher SD for inter-repetition interval compared with the NT group (91 ms vs. 22 ms). Inter-repetition intervals became shorter as the metronome rate became faster, as indicated by a strong effect of METRONOME SPEED ($t = -17.05$, $\beta = -.18$ $p < .001$, $R^2 = .81, .63$).

The ASD group showed a significantly higher *tap asynchrony* than the NT group ($t = 2.64$, $\beta = 52.69$, $p < .05$, $R^2 = .76, .21$) (Figure 7). The ASD group also showed a considerably higher standard deviation than the control group on this measure (126 ms vs. 8 ms). In addition, the ASD group showed significantly *higher inter-tap intervals* than the control group ($t = 3.08$, $\beta = 10.14$, $p < .01$, $R^2 = .22, .21$) (Figure 8); again, standard deviation for inter-tap interval was higher in the ASD group than the control group (21 ms vs. 2 ms). As predicted, inter-tap interval decreased with increasing METRONOME SPEED ($t = -67.48$, $\beta = -218.19$, $p < .001$, $R^2 = .98, .98$). Pearson's correlation revealed a moderately strong significant relationship between speech asynchrony and tap asynchrony across the tasks ($r = .49$, $p < .001$) (Figure 9).

[FIGURE 5 ABOUT HERE]

[FIGURE 6 ABOUT HERE]

[FIGURE 7 ABOUT HERE]

[FIGURE 8 ABOUT HERE]

[FIGURE 9 ABOUT HERE]

Effects of prosodic phrasing on timing

Results on prosodic phrasing on the speech cycling task generally matched our predictions (Table 2; Figure 10). For all models, a significant effect of PROSODIC FORM was found (Syllable 3: R^2 = .34, .03; Syllable 4: R^2 = .40, .02; Syllable 5: R^2 = .41, .01). As predicted, timing of **Syllable 3** (a non phrase-initial syllable) in the 3-3 grouping condition was overall earliest out of the five prosodic conditions. Syllable 3 occurred significantly later in the 2-4 grouping condition (t = 5.73, β =.01, $p < .001$), the 2-2-2 grouping condition (t = 10.63, β =.02, $p < .001$), the 4-2 grouping condition (t = 10.89, β =.02, $p < .001$), and the 2-3-1 grouping condition (t = 2.06, β =.00, $p < .05$), as compared with the 3-3 condition. For **Syllable 4**, the pattern changed, with Syllable 4 occurring earlier in the 2-2-2 grouping condition (t = -3.89, β =-.01, $p < .001$) (where it was a non phrase-initial syllable) than in the 3-3 condition. Syllable 4 also occurred quite late in the 2-1-3 grouping condition, even later than the 3-3 reference level (t = 3.52, β =.00, $p < .001$). Unexpectedly, no difference in timing was found for Syllable 4 between the 3-3 and 2-4 grouping conditions (t = 1.58, β =.00, $p = .11$). Also unexpectedly, Syllable 4 occurred significantly later in the 4-2 grouping condition (t = 7.25, β =.01, $p < .001$) compared with the 3-3 grouping condition. For **Syllable 5**, as expected, timing was significantly later in the 4-2 grouping condition compared with the 3-3 grouping condition (t = 7.60, β =.02, $p < .001$). No significant difference was found in timing of Syllable 5 from the reference level for the 2-4 condition (t = 1.90, β =.00, $p = .06$), the 2-2-2 condition (t = -.89, β =.00, $p = .37$) or the 2-1-3 condition (t = -.88, β =.00, $p = .38$). For Syllables 4 and 5, alignment position increased slightly with speech rate (positive ts > 3.45, βs =.01, ps < .01; Conditional R^2s > .40, Marginal R^2s > .01), while Syllable 3 remained more consistent in timing across the five speeds, as the effect of metronome speed was not significant (positive t = 1.80, β =.00, $p = .09$).

[TABLE 2 ABOUT HERE]

[FIGURE 10 ABOUT HERE]

Effects of Group on relative timing

Looking now to group effects on timing, we see that the ASD group showed overall earlier timing of syllables within the repetition cycle as compared with controls for all three syllables of interest (negative ts < -2.40, βs < -.02, ps < .01, Conditional R^2s > .30, Marginal R^2s > .06). This is notable, since it suggests that, even relative to their own, self-driven repetition cycles, the ASD group anticipated the timing of syllables in comparison with controls (recall that the timing of the metronome beat itself did not factor into relative phase measurements for the speech task). Results of relative timing are plotted for each syllable by group in Figure 11. On the whole, prosodic patterns were found to be similar across the two groups. However, significant interactions between GROUP and PROSODIC FORM indicate that there were some key differences between the groups. Specifically, there was a significant interaction in the comparison of the 3-3 prosodic condition with the 2-2-2 condition for all three syllables of interest. Recall that Syllable 3 occurred earlier overall in the 3-3 condition than in the 2-2-2, as predicted; this difference was actually found to be more extreme for the ASD group than in the NT group for the 2-2-2 comparison (t = 2.35, β =.01, $p < .05$). For Syllable 4, the expected difference found between the 3-3 and 2-2-2 grouping patterns was in fact *less* extreme in the ASD group than the NT group (t = 3.45, β =.01, $p < .001$). For Syllable 5, the difference in timing between the 3-3 and 2-2-2 group was again more strongly in the expected direction (with Syllable 5 in the 2-2-2 group slightly later than in the 3-3 group) for the ASD group than for the NT group (t = 2.85, β =.01, $p < .01$). Taken together, these results suggest that variation in relative timing patterns between the ASD and NT groups showed a more extreme level of prosodic phrasing by ASD participants in some cases (e.g. for Syllables 3 and 5) and a less extreme level in others (e.g. for Syllable 4).

[FIGURE 11 ABOUT HERE]

Effects of Group on syllable duration

There was a significant overall effect of SYLLABLE ORDER on duration, as predicted (R^2 = .24, .21): Syllable 6, the final syllable, was found to be significantly longer than medial Syllable 3 (t = 109.07, β =.08, $p < .001$). Differences were also found between Syllable 5, the penultimate syllable (t = 21.62, β =.02, $p < .001$), and Syllable 4, the antepenultimate syllable (t = 21.22, β =.02, $p < .001$), in comparison with Syllable 3, with duration increasing as syllables became closer to the phrase boundary. Syllable duration also predictably reduced as speech rate increased as a function of METRONOME SPEED (t = -12.92, β =.02, $p < .001$).

No overall difference was found between ASD and NT groups in syllable duration (t = -.26, β =.00, p = .80, R^2 =04, .002). However, significant interactions between GROUP and SYLLABLE ORDER emerged for comparisons between Syllable 3 and Syllable 6 (t = 10.61, β =.01, $p < .001$), Syllable 3 with Syllable 5 (t = -4.66, β =.00, $p < .001$), and Syllable 3 with Syllable 4 (t = -5.09, β =.00, $p < .001$). As can be seen in Figure 12, the ASD group exhibited less of a difference between Syllable 3 vs. Syllables 4 and 5, but a much larger difference between the duration of Syllable 3 and that of Syllable 6.

[FIGURE 12 ABOUT HERE]

Sensitivity and specificity of timing measures

All six of the timing variables examined significantly predicted group status (|z|s > 2.9, $|\beta|$s >.001, ps < .01). Of the six models, correct predictions for the ASD group (sensitivity) and for the control group (specificity) were highest for the tap asynchrony (96% and 68%, respectively) and speech asynchrony (90% and 52%) models. The inter-repetition and inter-tap interval models had similar rates of specificity (both around 52%), but both had lower sensitivity than the asynchrony models (68% for inter-repetition and 60% for inter-tap). The proportion model had a slightly higher rate of specificity (63%) but even lower sensitivity (43%). Finally, the duration model had a very high specificity rate (94%) but a very low sensitivity rate (5%). Note that these last two models took into consideration relative timing and duration of all syllables examined in the study, so were not specific to any one syllable.

Discussion

This study had two main goals: to investigate in detail patterns of prosodic phrasing among Cantonese-speaking individuals with ASD, and to explore the possibility that differences in prosodic patterning in the disorder could be linked with a more general timing deficit in ASD. Our results revealed several patterns of interest. First off, patterns of prosodic grouping, as indicated through relative timing on the speech cycling task, were found to be relatively similar across groups. Where differences emerged, for the most part, the ASD group in fact showed more prominent cues to prosodic grouping than the control group. An exception to this finding was for prosodic patterning on Syllable 4, where the ASD group failed to show the expected effects of grouping. We also observed more dramatic effects of prosodic lengthening on phrase-final syllables among the ASD group. This was accompanied by more moderate effects of pre-final lengthening on the penultimate and antepenultimate syllables compared with NT controls.

One key takeaway from our findings is therefore that certain timing distortions observed in ASD in languages like English —such as increased lengthening—are also present among Cantonese speakers with the disorder. However, our results also suggest that care must be taken in interpreting these patterns of increased lengthening across languages. Specifically, prosodic differences in ASD cannot be attributed solely to atypical speech rate or a global effect of syllable lengthening. Rather, our findings indicate that prosodic differences vary at different positions within a phrase, with syllables at phrase edges more likely to exhibit exaggerated prosodification, while those occurring closer to phrase-internal positions are prone to under-differentiation.

Differences in phrase-final lengthening in ASD

One possibility is that the observed asymmetries in duration on final versus pre-final syllables in the task reflect the relative greater salience of phrase-final vs. non-final syllables; speakers in the ASD group may have acquired a weaker degree of pre-final lengthening owing to this difference in perceptual salience. This would be consistent with evidence from language acquisition showing that children tend to faithfully produce prosodic patterns on phrase-final syllables, but prosodically reduce, or even delete, non-final syllables (Echols & Newport, 1992). However, this would not explain why the ASD group should additionally apply greater lengthening on the final syllable than the control group. The varied effects that the ASD group showed on final versus non-final syllables compared to controls could alternatively be explained through a difference in the application of the *clock-slowing* mechanism that has been proposed to drive final lengthening (Byrd & Saltzman, 2003). According to this theory, gradual changes in duration towards the phrase boundary occur as a result of an articulatory slowing gesture (π-gesture) whose activation level increases as the phrase boundary approaches. The direct effect of this clock slowing gesture is to reduce the stiffness of articulatory constriction gestures for consonants and vowels over which the π-gesture is active (Byrd, 2000). As the π-gesture reaches its peak activation at the phrase boundary, individual segment-level gestures (constrictions) decrease in stiffness and become longer and longer in duration. A schematization of the π-gesture is given in Figure 13.

[FIGURE 13 ABOUT HERE]

Byrd and Saltzman (2003) propose that there are three types of modulations which can be applied to the π-gesture to change how the resulting stiffness/durations of constriction gestures are affected: temporal alignment modulation, strength modulation, and shaping (skew) modulation. A temporal alignment modulation of the π-gesture will lead to an earlier or later onset of the gesture (Figure 14-b). A strength modulation will lead to a stronger or weaker peak activation of the gesture (Figure 14-c). Skewing of the activation level of the π-gesture will lead to further asymmetries in stiffness/lengthening across constriction gestures (Figure 14-d).

[FIGURE 14 ABOUT HERE]

Based on Figure 14, we can see that later temporal alignment and rightward skewing both lead to longer durations of later gestures than earlier ones; this is consistent with what is generally found with phrase-final lengthening. From this perspective, we could hypothesize that differences in durational patterns found in the ASD group stem from a difference in timing or shaping of the π-gesture wave, or perhaps a combination of both.

Intact relative timing, but poor metronome coordination

The ASD group showed overall similar patterns to the control group on prosodic phrasing as measured through relative timing despite the fact that their general timing within the speech repetition cycle was earlier, their coordination with the metronome consistently less accurate, and their individual syllable durations different from those of controls. This suggests that grammatical knowledge of prosodic grouping is intact in ASD, and that observed differences in prosodic patterning do not arise not from a grammatical deficit, but rather a deficit in speech motor planning and/or execution. Our findings also revealed, in line with previous work, that the ASD group showed less accuracy and greater variability in timing to the metronome for both the speech and drumming tasks. Eight out of the ten participants with the highest scores (=least accurate/most variable) on the asynchrony and variability measures were from the ASD group, suggesting that this is a consistent pattern across individuals with ASD. That diagnosis predicted timing difficulties (and vice versa) across speech and drumming tasks as well as observed differences in prosodic phrasing is consistent with the idea that a more general timing mechanism underlies all of these phenomena. Nevertheless, we note that large standard deviations in several of the accuracy and variability measures in the ASD group vs. the control group may mean that not all participants with ASD show a timing deficit to the same degree. A more in-depth study of individual differences in timing abilities in ASD will be useful for understanding these patterns and informing individualized clinical intervention strategies.

Subtypes of temporal processing and links with prosodic processing in ASD

Recent work in neuroscience suggest that there are two primary ways in which the brain tracks time: rhythmic timing, in which expectations about upcoming events are built based on a regularly-timed stimulus (regulated at least in part by the basal ganglia) and interval timing, which does not rely on the presence of a regularly-timed stimulus (regulated at least in part by the cerebellum) (Breska & Ivry, 2016). Our results accord with previous findings that individuals with ASD have trouble coordinating manual movements with a rhythmically-timed external stimulus (Gowen & Miall, 2005; Morimoto et al., 2018). This is consistent with the notion that individuals with ASD struggle with rhythmic timing; such difficulties have also been argued to underlie reduced patterns of neural entrainment to speech in ASD (Jochaut et al., 2015). An important caveat is that the present study only examined rhythmic timing in coordination with an external stimulus. There is some evidence to suggest that production of rhythmic timing is improved in ASD as long as participants may rely on their own, internally-generated rhythms, rather than coordinating with an external stimulus (Gowen & Miall, 2005). Therefore, it could be the coordinative nature of the two rhythmic tasks which caused individuals with ASD to perform more poorly on the tasks, rather than the rhythmic nature of the tasks, per se (see also Repp, 2002). However, we note that the two groups in our study also showed differences in relative timing within the self-generated speech repetition cycle, with the ASD group showing overall earlier timing within the cycle. This mirrors findings of overall greater tendency towards anticipation (rather than tracking or lagging) in the tapping task by the ASD group. Thus, we do have some evidence to suggest that rhythmic timing is impaired independently of coordination with the external stimulus. This finding furthermore suggests that deficits in rhythmic timing among ASD participants is not solely attributable to poor auditory perception or auditory-motor integration abilities.

The durational variability observed with respect to final lengthening and pre-final lengthening is interesting from the perspective of interval timing models, since these processes do not result in regularity of timing, and instead do quite the opposite: they stretch the timing of material at the end of the phrase with respect to what came before. Since final lengthening applies gradually to syllables at the end of an utterance, the process involves an intricate level of temporal planning on the part of the speaker in order to ensure lengthening begins at the right time and scales appropriately as the phrase boundary is reached. Final lengthening must also scale in proportion to speech rate and other factors (Berkovits, 1991). Previous work has shown that those with ASD have difficulty with timing in other goal-oriented motor tasks such as ball throwing and catching (Green et al., 2002, 2009); similar to our findings, it has been suggested that movement *slowing*, in particular, is impacted in complex movement tasks (Whyatt & Craig, 2013). Velocity and acceleration of controlled arm movements have been found to be overall greater in those with ASD compared with controls (Cook et al., 2013). Similarly, in their beat entrainment task, Morimoto et al. (2018) demonstrated that participants with ASD showed greater velocity in the closing movement of finger-thumb tapping movements. These findings all suggest that cognitive control of timing may be the source of distorted patterns of phrase-final lengthening in ASD. Additional kinematic studies will be useful to directly examine parallels in movement timing across speech and other motor domains like tapping and throwing.

Prosody and conversational timing deficits

Our findings may also have implications for understanding why conversational interaction is particularly difficult for those with ASD. It has been suggested that, in addition to difficulties making pragmatic inferences in conversation, individuals with ASD struggle to time their conversational turns appropriately (Paul et al., 2009). During a typical conversational interaction, conversation partners attend to patterns of final and prefinal lengthening in their partner's speech at turn ends so as to begin to plan their own next turn (Jefferson, 1986). If durational cues at turn ends are not implemented as expected so that partners can make predictions about where to begin their turns, conversation may become more stilted. Conversely, if individuals with ASD are less able to process turn boundary cues, they may initiate a turn in a way that feels rhythmically awkward to their conversation partner in the context of the conversation.

Clinical Implications

Our results indicate that individuals with ASD struggle with speech timing patterns associated with organizational aspects of prosody, and not just social-pragmatic aspects of prosody use. This suggests that clinical interventions which focus on helping individuals to regulate timing—such as auditory rhythmic cuing—may be especially useful for treating disordered prosody in ASD. Such therapies have been found to be successful in helping those with ASD to enhance timing abilities in other motor domains (El Shemy & El-Sayed, 2018). Though it has been suggested that such therapies might also be useful for improving prosody and interpersonal communicative timing (Hardy & LaGasse, 2013), the efficacy of such treatments has yet to be investigated; collaboration between researchers and clinicians in this area will be a fruitful avenue for future work. It is also notable that rhythmic asynchronies on both the speech and tapping tasks were found to be highly sensitive measures of autism diagnosis with relatively high levels of specificity. Given the ease of implementation of the speech cycling task, this task may serve as a useful tool in assessing speech and motor deficits in ASD.

Study Limitations and Future Directions

The present study has some important limitations. For one, the study did not incorporate an un-paced condition on either the speech or tapping tasks. We are therefore unable to assess the degree to which the coordinative nature of the metronome task may have influenced our results. It is possible, for example, that prosodic differences observed between our two groups would not have arisen (or may have been weaker) outside the constraints of the metronome task. We note, however, even in the context of naturalistic speech, final and pre-final lengthening processes require speakers to meter out time during the course of an utterance, applying lengthening proportionally to their speech rate and to the size of the prosodic unit (Byrd & Salzman, 2003). Generally speaking, results from the speech cycling task also mirror speech prosody patterns found in naturalistic speech (Cummins & Port, 1998). We therefore have good reason to think that the timing patterns observed among participants in the current study are representative of their prosodic timing patterns more broadly.

We also cannot conclude, based on the present results, the precise nature of temporal processing in ASD and how it relates to speech prosody. Both the speech cycling task and the drumming task used in the present study likely involve several types of temporal processing, including those implicated in movement timing, temporal prediction, feedback-based learning, and internal modeling of sensorimotor dynamics (see Repp, 2005 for an overview). To delve deeper into the mechanism for temporal processing in speech in ASD, it will be necessary to employ tasks which are more circumscribed in their role for assessing specific temporal difficulties, such as the temporal bisection task.

Finally, we take care to point out that the results of the present study, while suggestive, are not conclusive in motivating an ontogenetic causal relationship between temporal processing and speech prosody deficits in ASD. It is possible that temporal processing deficits are simply an epiphenomenon of the disorder and do not play a causal role. Further research into developmental trajectories of temporal processing and language in autism spectrum disorders will be necessary in order to elucidate the precise causal chain.

References

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences*, *12*(10), 366–375. https://doi.org/10.1016/0166-2236(89)90074-X

Alcantara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

Allman, M. J. and De Leon I. G. (2009). No time like the present: Time perception in autism. In Giordano, A. C. & Lombardi, V. A. (Eds.), *Causes and Risks for Autism* (pp. 65-76). New York: Nova Science.

Allman, M. J. (2011). Deficits in Temporal Processing Associated with Autistic Disorder. *Frontiers in Integrative Neuroscience*, *5*. https://doi.org/10.3389/fnint.2011.00002

Allman, M. J., DeLeon, I. G., & Wearden, J. H. (2011). Psychophysical Assessment of Timing in Individuals With Autism. *American Journal on Intellectual and Developmental Disabilities*, *116*(2), 165–178. https://doi.org/10.1352/1944-7558-116.2.165

Allman, M., & Falter, C. (2015). Abnormal Timing and Time Perception in Autism Spectrum Disorder?: A Review of the Evidence. In Vatakis A. & Allman M. (Eds.), *Time*

Distortions in Mind: Temporal Processing in Clinical Populations (pp. 37-56). Leiden; Boston: Brill. Retrieved June 2, 2020, from www.jstor.org/stable/10.1163/j.ctt1w8h2wk.7

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

American Speech-Language-Hearing Association. (2020). Guidelines for manual pure-tone threshold audiometry. [Guidelines]. Available from https://www.asha.org/policy/GL2005-00014/

Asperger, H. (1952). *Healpädagogik*. Berlin: Springer.

Asperger, H., & Frith, U. (1991). "Autistic psychopathy" in childhood. In *Autism and Asperger syndrome* (pp. 37–92). Cambridge: Cambridge University Press.

Baddeley, A. (1986). *Working memory.* (pp. xi, 289). Clarendon Press/Oxford University Press.

Baddeley, A. D. (1990). *Human memory: Theory and practice.* (pp. xi, 515). Allyn & Bacon.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., & Staunton, R. (1994). Do children with autism acquire the phonology of their peers? An examination of group identification through the window of bilingualism. *First Language*, *14*(42–43), 241–248.

Barr, D., Levy. R., Scheepers, C., and Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255–278.

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech* (pp. 152–178). Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309. https://doi.org/10.1017/S095267570000066X

Berger, D. S. (2012). Pilot study investigatingn the efficacy of tempo-specific rhythm interventions in music-based treatment addressing hyper-arousal, anxiety, system pacinng, and redirection of fight-or-flight fear behaviors in children with Autism Spectrum Disorder (ASD). *Journal of Biomusical Engineering*, *2*, 1–15.

Berkovits, R. (1991). The Effect of Speaking Rate on Evidence for Utterance-Final Lengthening. *Phonetica*, *48*(1), 57–66. https://doi.org/10.1159/000261871

Bhat, A. N., Landa, R. J., & Galloway, J. C. (Cole). (2011). Current Perspectives on Motor Functioning in Infants, Children, and Adults With Autism Spectrum Disorders. *Physical Therapy*, *91*(7), 1116–1129. https://doi.org/10.2522/ptj.20100294

Boersma, P. & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.37, retrieved March 2018 from http://www.praat.org/

Breska, A., & Ivry, R. B. (2018). Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proceedings of the National Academy of Sciences*, *115*(48), 12283–12288. https://doi.org/10.1073/pnas.1810596115

Byrd, D. (2000). Articulatory Vowel Lengthening and Coordination at Phrasal Junctures. *Phonetica*, *57*(1), 3–16. https://doi.org/10.1159/000028456

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149–180. https://doi.org/10.1016/S0095-4470(02)00085-2

Van Bourgondien, M. E., & Woods, A. V. (1992). Vocational Possibilities for High-Functioning Adults with Autism. In E. Schopler & G. B. Mesibov (Eds.), *High-Functioning Individuals with Autism* (pp. 227–239). Springer US. https://doi.org/10.1007/978-1-4899-2456-8_12

Brown, L., & Johnson, S. K. (2010). *Test of Nonverbal Intelligence*. Auxtin, TX: PRO-ED.

Van Buren, J. M. (1963). Confusion and Disturbance of Speech from Stimulation in Vicinity of the Head of the Caudate Nucleus. *Journal of Neurosurgery*, *20*(2), 148–157. https://doi.org/10.3171/jns.1963.20.2.0148

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173–199. https://doi.org/10.1006/jpho.1998.0071

Chan, K. K. L., & To, Carol K. S. (2016). Do individuals with high-functioning autism who speak a tone language show intonation deficits? *Journal of Autism and Developmental Disorders*, *46*, 1784–1792.

Chow, I., Brown, S., Poon, M., & Weishaar, K. (2010). A musical template for phrasal rhythm in spoken Cantonese. In *Speech Prosody 2010-Fifth International Conference.*

Chung, Y., & Arvaniti, A. (2013). Speech rhythm in Korean: Experiments in speech cycling. In *Proceedings of Meetings on Acoustics, Montreal 2-7 June, 19,* 060216, NY: Acoustic Society of America.

Cook, J. L., Blakemore, S.-J., & Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. *Brain*, *136*(9), 2816–2824. https://doi.org/10.1093/brain/awt208

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, *45*(1), 119–130. https://doi.org/10.1016/j.cortex.2007.09.008

Cummins, F. (1997). *Rhythmic Coordination in English Speech: An Experimental Study* [PhD Thesis]. Indiana University.

Cummins, Fred. (1999). Synergetic Organization in Speech Rhythm. In W. Tschacher & J.-P. Dauwalder (Eds.), *Dynamics, Synergetics, Autonomous Agents: Nonlinear systems approaches to Cognitive Psychology and Cognitive Science. Studies of Nonlinear Phenomena in Life Science – Volume 8* (pp. 256-266). New Jersey : World Scientific.

Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress-timing in English. *Journal of Phonetics*, *31*, 139-148. Dauer, R. M. (1983) Stress-timing and syllable-timing re-analysed. *Journal of Phonetics*, *11*, 51–62.

DePriest, J., Glushko, A., Steinhauer, K., & Koelsch, S. (2017). Language and music phrase boundary processing in Autism Spectrum Disorder: An ERP study. *Scientific Reports*, *7*(1), 14465. https://doi.org/10.1038/s41598-017-14538-y

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*(2), 144–152. https://doi.org/10.1016/j.bandl.2008.04.002

Diehl, J. J., & Berkovits, L. D. (2010). Is prosody a diagnostic and cognitive bellwhether of autism spectrum disorders? In A.E. Harrison (Ed.), *Speech Disorders: Causes, Treatments and Social Effects* (pp. 159–176). Nova Science Publishers, Inc.

Diehl, J. J., & Paul, R. (2009). The assessment and treatment of prosodic disorders and neurological theories of prosody. *International Journal of Speech-Language Pathology*, *11*(4), 287–292. https://doi.org/10.1080/17549500902971887

Echols, C. H., & Newport, E. L. (1992). The Role of Stress and Position in Determining First Words. *Language Acquisition*, *2*(3), 189–220. https://doi.org/10.1207/s15327817la0203_1

El Shemy, S. A., & El-Sayed, M. S. (2018). The impact of auditory rhythmic cueing on gross motor skills in children with autism. *Journal of Physical Therapy Science*, *30*(8), 1063–1068. https://doi.org/10.1589/jpts.30.1063

Estes, A., Shaw, D. W. W., Sparks, B. F., Friedman, S., Giedd, J. N., Dawson, G., Bryan, M., & Dager, S. R. (2011). Basal ganglia morphometry and repetitive behavior in young children with autism spectrum disorder. *Autism Research*, *4*(3), 212–220. https://doi.org/10.1002/aur.193

van der Fels, I. M. J., te Wierike, S. C. M., Hartman, E., Elferink-Gemser, M. T., Smith, J., & Visscher, C. (2015). The relationship between motor skills and cognitive skills in 4–16 year old typically developing children: A systematic review. *Journal of Science and Medicine in Sport*, *18*(6), 697–703. https://doi.org/10.1016/j.jsams.2014.09.007

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*(2), 233–253.

Fon, J., Johnson, K., & Chen, S. (2011). Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech. *Language and Speech*, *54*(1), 5–32. https://doi.org/10.1177/0023830910372492

Fosnot, S. M., & Jun, S.-A. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. *14th International Congress of Phonetic Sciences*, 103–115.

Franich, K. H. (2018). *The Interaction of Prominence, Rhythm, and Tone in Medumba*. https://doi.org/10.6082/M14X55ZW

Franich, K. (2019). Uncovering Tonal and Temporal Correlates of Phrasal Prominence in Medʉmba. *Language and Speech*, 002383091988799. https://doi.org/10.1177/0023830919887994

Frith, U. (Ed.). 1991. Autism and Asperger syndrome. Cambridge: Cambridge University Press.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*(4), 765–768.

Gowen, E., & Miall, R. C. (2005). Behavioural aspects of cerebellar function in adults with Asperger syndrome. *The Cerebellum*, *4*(4), 279–289. https://doi.org/10.1080/14734220500355332

Green, D., Baird, G., Barnett, A. L., Henderson, L., Huber, J., & Henderson, S. E. (2002). The severity and nature of motor impairment in Asperger's syndrome: A comparison with Specific Developmental Disorder of Motor Function. *Journal of Child Psychology and Psychiatry*, *43*(5), 655–668. https://doi.org/10.1111/1469-7610.00054

Green, D., Charman, T., Pickles, A., Chandler, S., Loucas, T., Simonoff, E., & Baird, G. (2009). Impairment in movement skills of children with autistic spectrum disorders. *Developmental Medicine & Child Neurology*, *51*(4), 311–316. https://doi.org/10.1111/j.1469-8749.2008.03242.x

Groen, W. B., van Orsouw, L., Huurne, N. ter, Swinkels, S., van der Gaag, R.-J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact Spectral but Abnormal Temporal Processing of Auditory Stimuli in Autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750. https://doi.org/10.1007/s10803-008-0682-3

Grosjean, F., Lane, H., Battison, R., & Teuber, H. (1981). The invariance of sentence performance structures across language modality. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 216–230. https://doi.org/10.1037/0096-1523.7.1.216

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and Affective Prosody in Children With High-Functioning Autism. *Journal of Speech, Language, and Hearing Research*, *53*(3), 778–793. https://doi.org/10.1044/1092-4388(2009/08-0127)

Hallett, M., Lebiedowska, M. K., Thomas, S. L., Stanhope, S. J., Denckla, M. B., & Rumsey, J. (1993). Locomotion of Autistic Adults. *Archives of Neurology*, *50*(12), 1304–1308. https://doi.org/10.1001/archneur.1993.00540120019007

Hardy, M. W., & LaGasse, A. B. (2013). Rhythm, movement, and autism: Using rhythmic rehabilitation research as a model for autism. *Frontiers in Integrative Neuroscience*, *7*. https://doi.org/10.3389/fnint.2013.00019

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, *38*(2), 241–253. https://doi.org/10.1016/0022-0965(84)90124-3

Jefferson, G. (1986). Notes on 'latency' In overlap onset. *Human Studies*, *9*(2–3), 153–183. https://doi.org/10.1007/BF00148125

Jobbágy, Á., Harcos, P., Karoly, R., & Fazekas, G. (2005). Analysis of finger-tapping movement. *Journal of Neuroscience Methods*, *141*(1), 29–39. https://doi.org/10.1016/j.jneumeth.2004.05.009

Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A.-D., Olasagasti, I., Chabane, N., Zilbovicius, M., & Giraud, A.-L. (2015). Atypical coordination of cortical oscillations in response to speech in autism. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00171

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *35*(2), 217–250.

Kanner, L. (1968). Autistic disturbances of affective contact. *Acta Paedopsychiatrica*, *35*(4), 100–136.

Kaur, M., Srinivasan, S. M., & Bhat, A.N. (2018). Comparing motor performance, praxis, coordination, and innterpersonal synchrony between children with annd without Autism Spectrum Disorder (ASD). *Research in Developmental Disabilities*, *72*, 79–95.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*(2), 162–179. https://doi.org/10.1016/j.wocn.2006.04.001

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Langen, M., Leemans, A., Johnston, P., Ecker, C., Daly, E., Murphy, C. M., dell'Acqua, F., Durston, S., & Murphy, D. G. M. (2012). Fronto-striatal circuitry and inhibitory control in autism: Findings from diffusion tensor imaging tractography. *Cortex*, *48*(2), 183–193. https://doi.org/10.1016/j.cortex.2011.05.018

Le Couteur, A., Lord, C., & Rutter, M. (2003). *Autism diagnostic interview-revised.* Los Angeles, CA: Western Psychological Services.

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1–2), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052

Lepistö, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Näätänen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neurophysiology*, *117*(10), 2161–2171. https://doi.org/10.1016/j.clinph.2006.06.709

Li, B. (2006). *Chinese final particles and the syntax of the periphery* [PhD Thesis]. Leiden University.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336.

Lin, H.-Y., & Fon, J. (2009). Perception of Temporal Cues at Discourse Boundaries. *Proceedings of INTERSPEECH*, 808–811.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule–2nd edition.* Los Angeles, CA: Western Psychological Corporation.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350.

Martin, J. S., Poirier, M., & Bowler, D. M. (2010). Brief Report: Impaired Temporal Reproduction Performance in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *40*(5), 640–646. https://doi.org/10.1007/s10803-009-0904-3

MATLAB and Statistics Toolbox Release 2012b. (2012). The MathWorks, Inc., Natick, Massachusetts, United States.

Mesibov, G. B. (1992). Treatment Issues with High-Functioning Adolescents and Adults with Autism. In E. Schopler & G. B. Mesibov (Eds.), *High-Functioning Individuals with Autism* (pp. 143–155). Springer US. https://doi.org/10.1007/978-1-4899-2456-8_8

Miendlarzewska, E. A., & Trost, W. J. (2014). How musical training affects cognitive development: Rhythm, reward and other modulating variables. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00279

Morimoto, C., Hida, E., Shima, K., & Okamura, H. (2018). Temporal Processing Instability with Millisecond Accuracy is a Cardinal Feature of Sensorimotor Impairments in Autism Spectrum Disorder: Analysis Using the Synchronized Finger-Tapping Task. *Journal of*

Autism and Developmental Disorders, *48*(2), 351–360. https://doi.org/10.1007/s10803-017-3334-7

Nadig, A., & Shaw, H. (2012). Acoustic and Perceptual Measurement of Expressive Prosody in High-Functioning Autism: Increased Pitch Range and What it Means to Listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dodrecht: Foris Publications.

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *The Journal of Neuroscience*, *37*(38), 9249–9258. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology*, *15*(9), e1007321. https://doi.org/10.1371/journal.pcbi.1007321

Patel, S. P., Nayar, K., Martin, G. E., Franich, K., Crawford, S., Diehl, J. J., & Losh, M. (2020). An Acoustic Characterization of Prosodic Differences in Autism Spectrum Disorder and First-Degree Relatives. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-020-04392-9

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124. https://doi.org/10.1016/j.rasd.2007.04.001

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational Behaviors in Youth with High-functioning ASD and Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *39*(1), 115–125. https://doi.org/10.1007/s10803-008-0607-1

Pickett, E. R., Kuniholm, E., Protopapas, A., Friedman, J., & Lieberman, P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: A case study. *Neuropsychologia*, *36*(2), 173–188. https://doi.org/10.1016/S0028-3932(97)00065-1

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [PhD Thesis]. MIT.

Port, R., Tajima, K., & Cummins, F. (1999). Speech and rhythmic behavior. In G. J. P. Savelsburgh, H., van der Maas, & P. C. L. van Geert (Eds.), *The non-linear analysis of developmental processes*. (pp.1-19). Amsterdam: Elsevier.

R Core Team (2013). R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. https://doi.org/10.3758/BF03206433

Rinehart, N. J., Tonge, B. J., Bradshaw, J. L., Iansek, R., Enticott, P. G., & Johnson, K. A. (2006). Movement-related potentials in high-functioning autism and Asperger's disorder. *Developmental Medicine & Child Neurology*, *48*(04), 272. https://doi.org/10.1017/S0012162206000594

Rogers, T. D., Dickson, P. E., McKimm, E., Heck, D. H., Goldowitz, D., Blaha, C. D., & Mittleman, G. (2013). Reorganization of Circuits Underlying Cerebellar Modulation of Prefrontal Cortical Dopamine in Mouse Models of Autism Spectrum Disorder. *The Cerebellum*, *12*(4), 547–556. https://doi.org/10.1007/s12311-013-0462-2

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.

Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., Tan, M., Nadig, A., & Hyde, K. (2018). Music improves social communication and auditory-motor connectivity in children with autism. *Translational psychiatry*, *8*(1), 231. https://doi.org/10.1038/s41398-018-0287-3.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193–247.

Srinivasan, S. M., Park, I. K., Neelly, L. B., & Bhat, A. N. (2015). A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with Autism Spectrum Disorder (ASD). *Research in Autism Spectrum Disorders*, *18*, 51–63. https://doi.org/10.1016/j.rasd.2015.07.004

Sussman, D., Leung, R. C., Vogan, V. M., Lee, W., Trelle, S., Lin, S., Cassel, D. B., Chakravarty, M. M., Lerch, J. P., Anagnostou, E., & Taylor, M. J. (2015). The autism puzzle: Diffuse but not pervasive neuroanatomical abnormalities in children with ASD. *NeuroImage: Clinical*, *8*, 170–179. https://doi.org/10.1016/j.nicl.2015.04.008

Szelag, E., Kowalska, J., Galkowski, T., & Pöppel, E. (2004). Temporal processing deficits in high-functioning children with autism. *British Journal of Psychology*, *95*(3), 269–282. https://doi.org/10.1348/0007126041528167

Tajima, K. (1998). *Speech rhythm in English and Japanese: Experiments in Speech Cycling* [PhD Thesis]. Indiana University.

Tajima, K., & Port, R. F. (2003). Speech rhythm in English and Japanese. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI*. (pp. 322-339). Cambridge: Cambridge University Press.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, *5*(11), 1226–1235. https://doi.org/10.1038/nn963

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Vernazza-Martin, S., Martin, N., Vernazza, A., Lepellec-Muller, A., Rufo, M., Massion, J., & Assaiante, C. (2005). Goal Directed Locomotion and Balance Control in Autistic Children. *Journal of Autism and Developmental Disorders*, *35*(1), 91–102. https://doi.org/10.1007/s10803-004-1037-3

Whyatt, C., & Craig, C. (2013). Sensory-motor problems in Autism. *Frontiers in Integrative Neuroscience*, *7*. https://doi.org/10.3389/fnint.2013.00051

Wong, W. P., Chan, M. K. M., & Beckman, M. E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. In S.-A. Jun (Ed.), *Prosodic Typology* (2nd ed.). Oxford University Press.

Yip, S. (2018) What determines Cantonese speech rhythm? [Master's Thesis]. Hong Kong University.

Appendix 1: Full list of experimental stimuli

noun//verb/noun//verb/noun 222	222 NP [V N] [V N] 家姐開燈溫書 gaa1ze1// hoi1/dang1//wan1/syu1 Sister switched on the light to work. noun//verb/noun//verb/noun
	222 NP [V N] [V N] 先生迫車返工 sin1saang1// bik1/ce1// faan1/gung1 Sir being crowded in transport on the way to work noun//verb/noun//verb/noun
	222 NP [V N] [V N] 孫姨熄燈休息 syn1ji1// sik1/dang1//jau1/sik1 Auntie Suen switched off the light and get rest. noun//verb/noun//verb/noun
noun//verb//noun 213	213 NP V NP 姑媽返烏溪沙 gu1maa1//faan1//wu1kai1sa1 Auntie goes back to Wu Kai Sha. noun//verb//noun
	213 NP V NP 叮噹聽收音機 ding1dong1//teng1// sau1jam1gei1 Shan Shan listens to the radio noun//verb//noun
	213 NP V NP 珊珊煎叉燒包 saan1saan1//zin1//caa1siu1/baau1 Doreamon pan fried roasted pork bun. noun//verb//adj/noun
verb/aspest-marker//noun213	V-ASP NP 執返啲朱古力 zap1/faan1//di1/zyu1gu1lik1 Re-collecting the chocolate

	verb/aspest-marker//noun
	V NP 分開堆污糟衫 fan1/hoi1//deoi1/wu1zou1/saam1 Separate the dirty clothes verb/verb-particle//adj/adj/noun
	V-ASP NP 拎開枝燒烤叉 ling1/hoi1/zi1/syu1/haau1/caa1 Take away the barcecue fork
verb//noun//verb//noun	33 [V NP] [V NP] 執支筆返公司 zap1//zi1/bat1//faan1//gung1si1 Pick a pencil to office verb//adj/noun//verb//noun
	33 [V NP] [V NP] 拎膠叉拮香蕉 ling1//gaau1/caa1//gat1/hoeng1ziu1 Take plastic fork to eat banana verb//adj/noun//verb//noun
	33 [V NP] [V NP] 煲雞湯加花生 bo1/gai1/tong1/gaa1/faa1/saang1/ Add fish maw for boiling chicken soup
noun//verb 42	42 [NP] [V] 骨科醫生返工 guat7/fo1/ji1/sang1/faan1/gung1 Orthopedist goes to work
	42 [NP] [V] 司機先生查車 si1/gei1/sin1/saang1/za1/ce1 Driver drives the car
	42 [NP] [V] 先施公司開倉 sin1/si1/gung1/si1/hoi1/cong1/ Sincere Department Store is on sale

Figure Captions

Figure 1. An example of a six-syllable English sentence as it is typically timed with a metronome in the speech cycling task. It has often been found that the last stressed syllable of the phrase (underlined here) preferentially aligns with lower-order fractions of the repetition cycle, such as the halfway point.

Figure 2. Predicted alignment differences across syllables residing in different prosodic positions in the speech cycling task (boundaries of prosodic phrases indicated with brackets

Figure 3. Grouping of syllables into hierarchically-organized prosodic constituents in Cantonese

Figure 4. Calculation of relative syllable timing for σ6 is interval *b* divided by interval *a*

Figure 5. Group effects on speech asynchrony. Error bars and ribbons reflect 95% confidence intervals.

Figure 6. Group effects on speech repetition interval

Figure 7. Group effects on tap asynchrony

Figure 8. Group effects on inter-tap interval

Figure 9. Relationship between tap asynchrony and speech asynchrony (averaged by metronome speed)

Figure 10. Relative timing by prosodic form condition

Figure 11. Relative timing by group and prosodic form condition

Figure 12. Syllable duration by syllable and group

Figure 13. Application of an intonational phrase-final π-gesture to consonant (C) and vowel (V) constriction gestures for the final 4 syllables of a 6-syllable utterance. Activation level of the gesture reaches its peak (darkest grey portion) during the constriction for the phrase-final syllable.

Figure 14. Adapted from Byrd & Saltzman (2003): effects of later timing (c), greater strength (d), and positive skew shaping (e) modulations on π-gesture activation patterns across two consecutive articulatory constriction gestures in comparison with unmodulated gestures (a) and a centered, unskewed π-gesture with relatively weak strength (b). The duration of later gestures is increased relative to earlier gestures in (c) and (e).

Figure 1

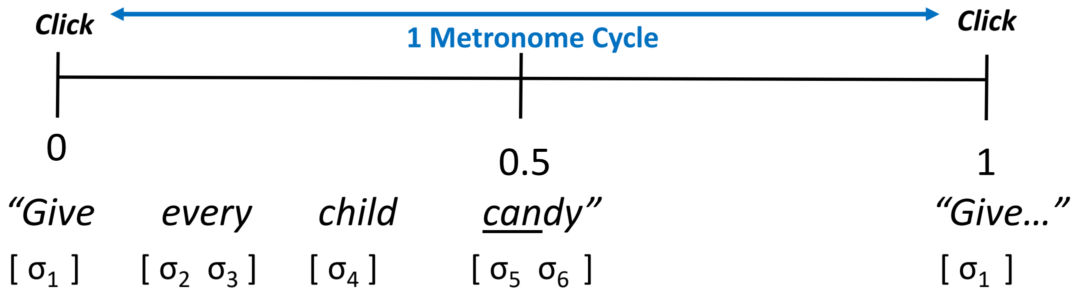

Click
1 Metronome Cycle
Click
0
0.5
1
"Give every child candy"
[σ1] [σ2 σ3] [σ4] [σ5 σ6]
"Give…"
[σ1]

Figure 2

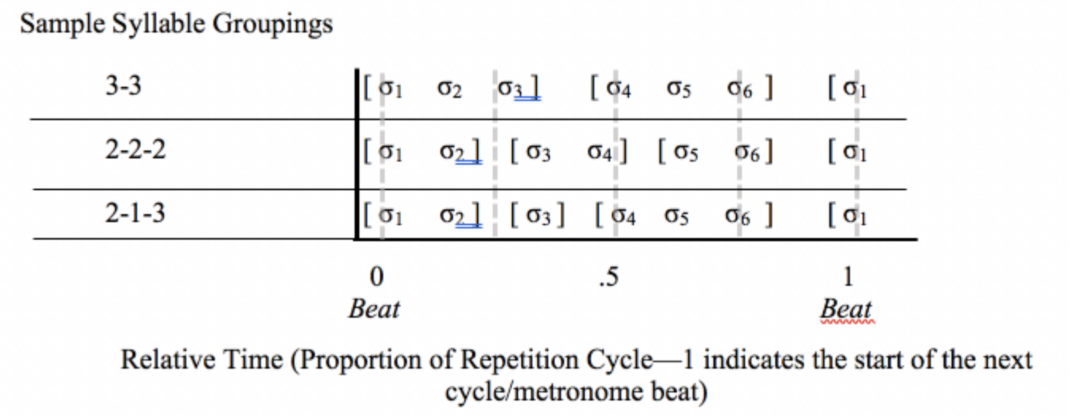
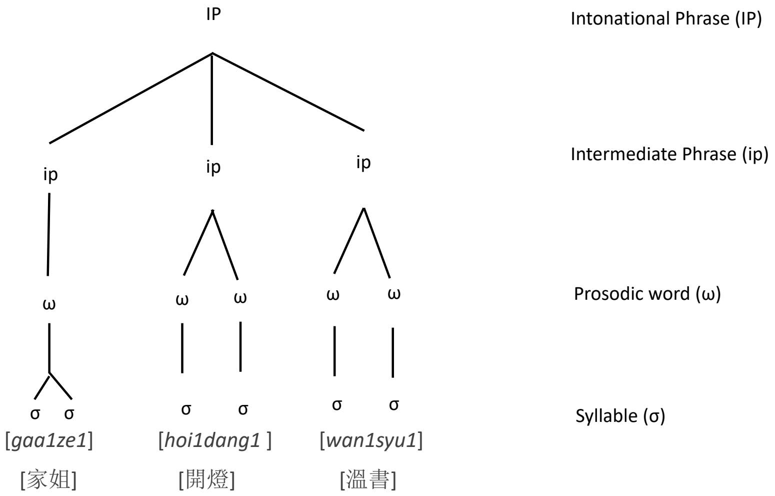

Sample Syllable Groupings
3-3
[σ_1 σ_2 σ_3] [σ_4 σ_5 σ_6] [σ_1
2-2-2
[σ_1 σ_2] [σ_3 σ_4] [σ_5 σ_6] [σ_1
2-1-3
[σ_1 σ_2] [σ_3] [σ_4 σ_5 σ_6] [σ_1
0
Beat
.5
1
Beat
Relative Time (Proportion of Repetition Cycle—1 indicates the start of the next cycle/metronome beat)

Figure 3



“Sister switches on the light to study.”

Figure 4

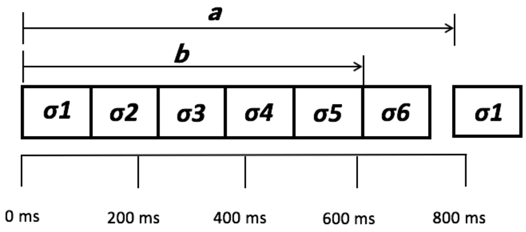


Figure 5

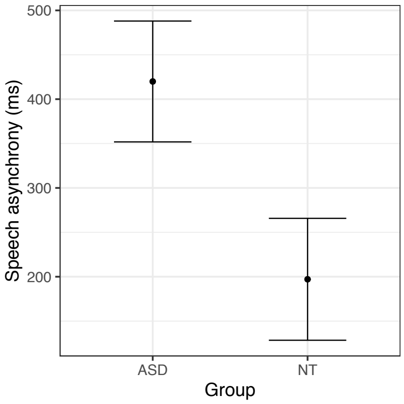


Figure 6

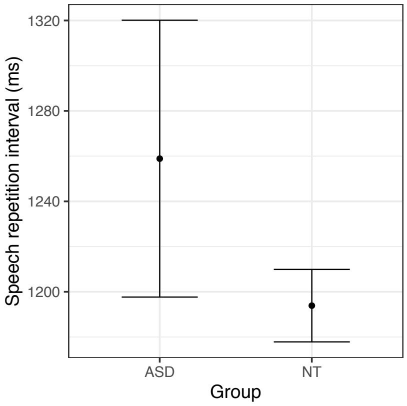

Speech repetition interval (ms)
1320
1280
1240
1200
ASD
NT
Group

Figure 7

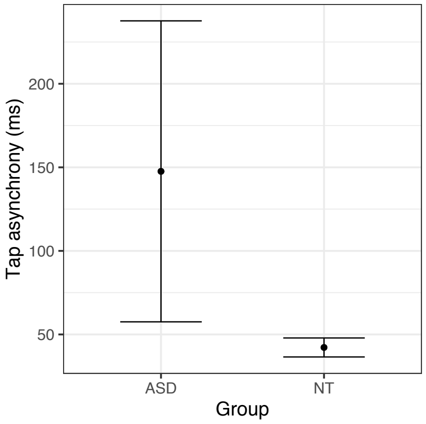


Figure 8

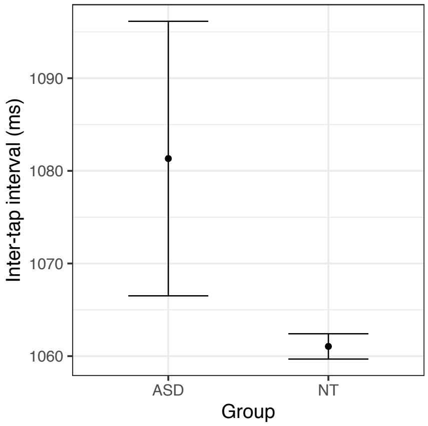


Figure 9

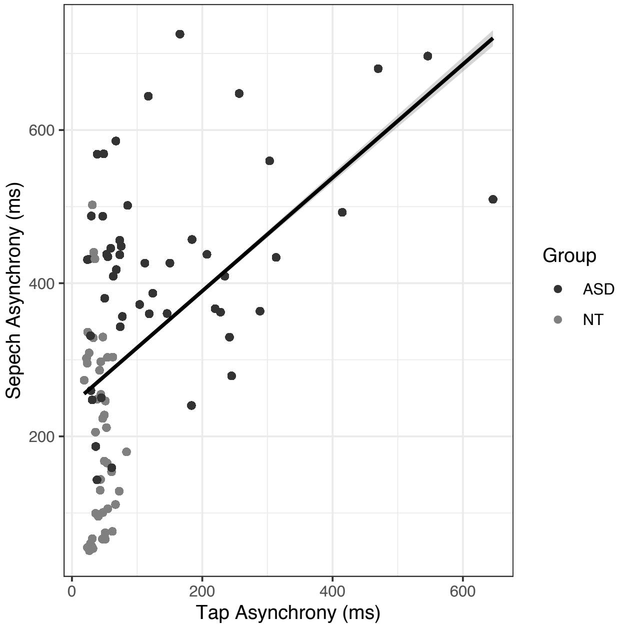

Sepech Asynchrony (ms)
600
400
200
0
200
400
600
Tap Asynchrony (ms)
Group
● ASD
● NT

Figure 10

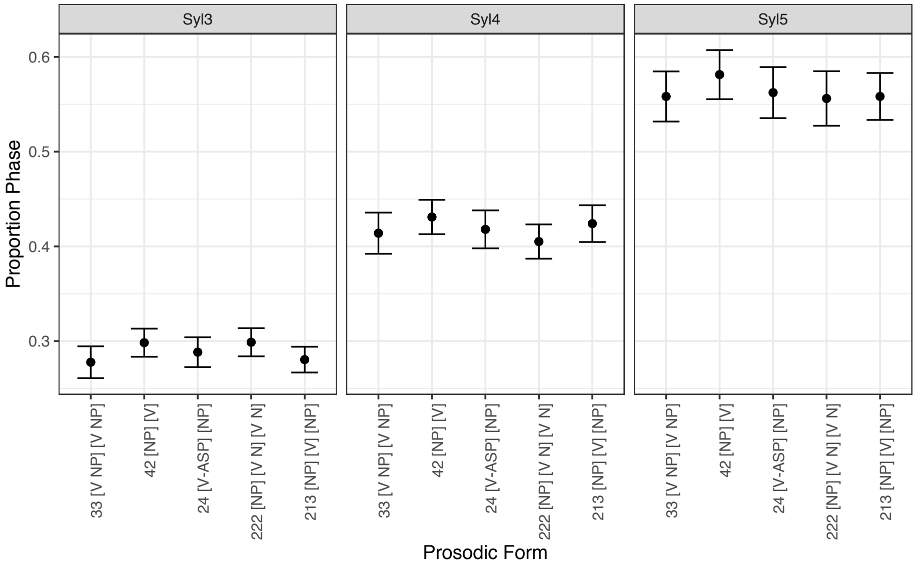

Syl3
Syl4
Syl5
Proportion Phase
0.6
0.5
0.4
0.3
33 [V NP] [V NP]
42 [NP] [V]
24 [V-ASP] [NP]
222 [NP] [V N] [V N]
213 [NP] [V] [NP]
Prosodic Form

Figure 11

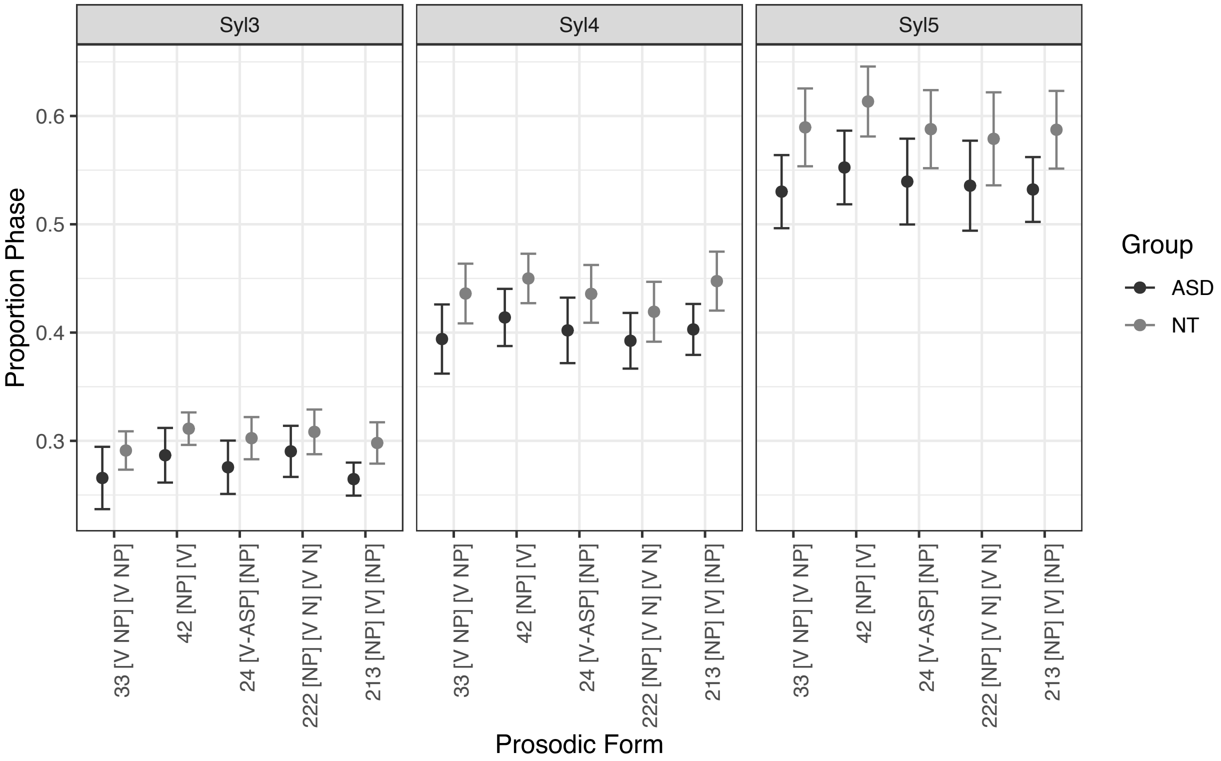

Syl3
Syl4
Syl5
Proportion Phase
0.6
0.5
0.4
0.3
33 [V NP] [V NP]
42 [NP] [V]
24 [V-ASP] [NP]
222 [NP] [V N] [V N]
213 [NP] [V] [NP]
Prosodic Form
Group
ASD
NT

Figure 12

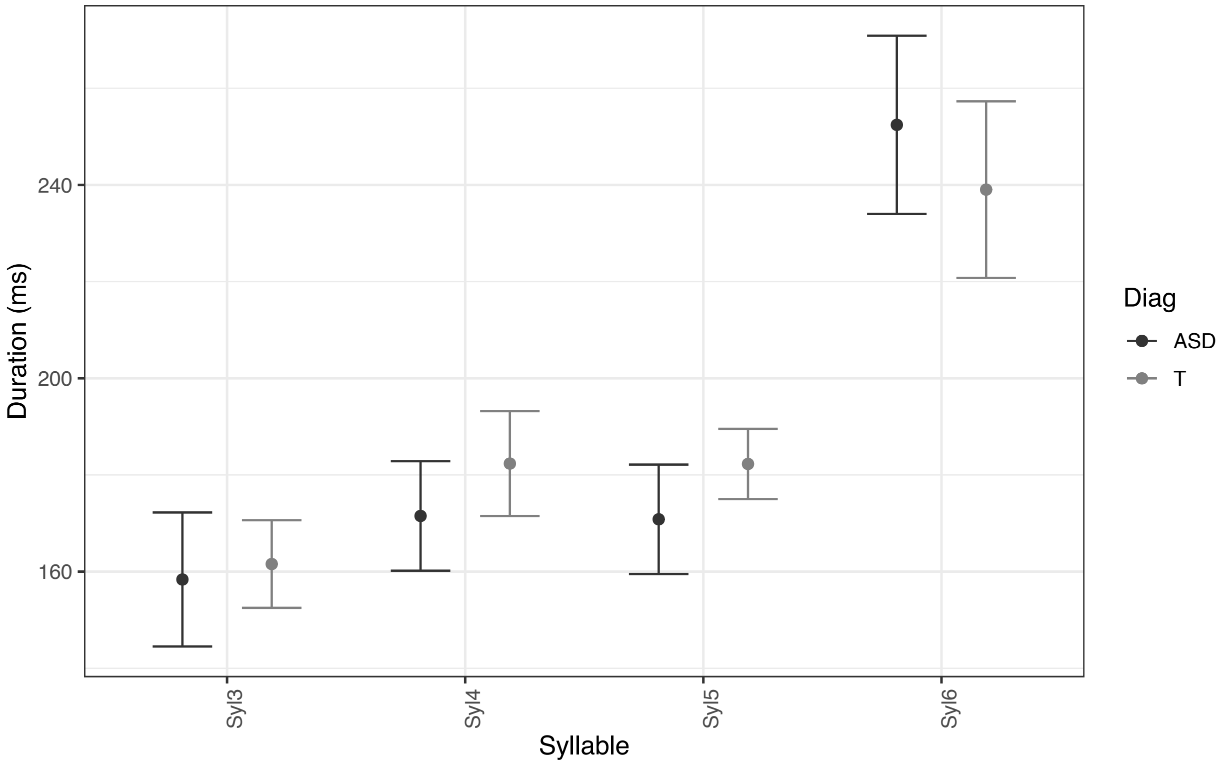

Duration (ms)
240
200
160
Syl3
Syl4
Syl5
Syl6
Syllable
Diag
ASD
T

Figure 13

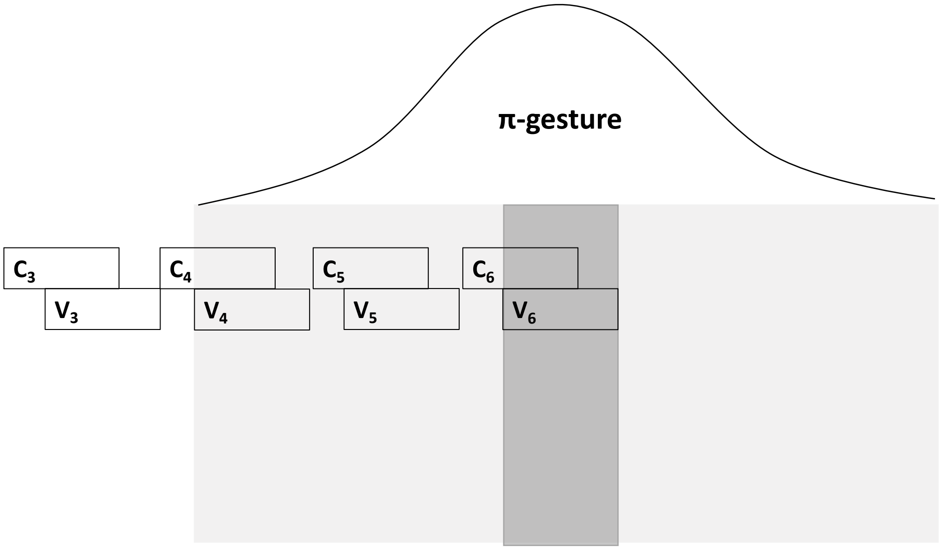


Figure 14 (a)

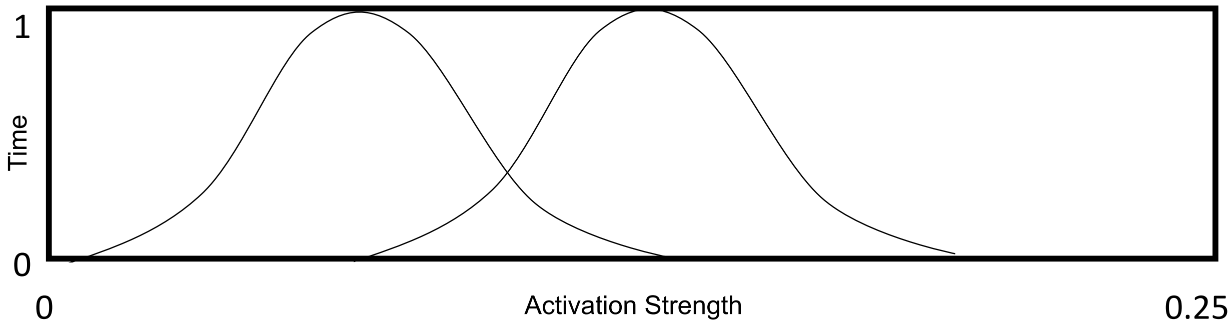


Figure 14 (b)

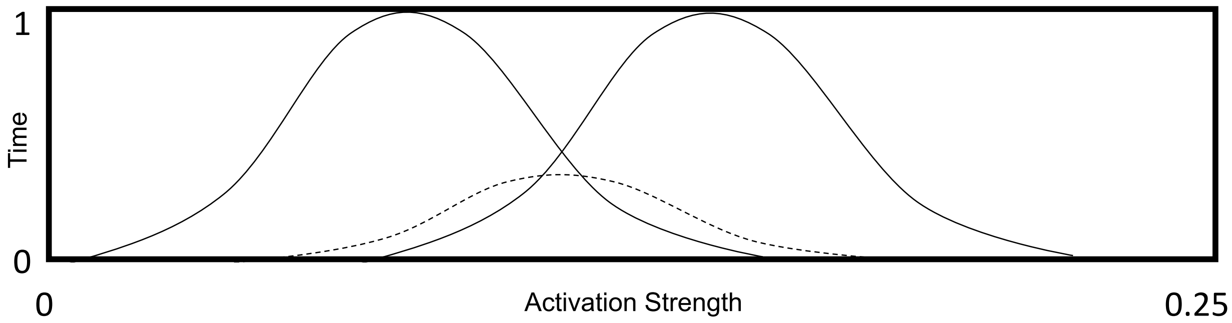


Figure 14 (c)

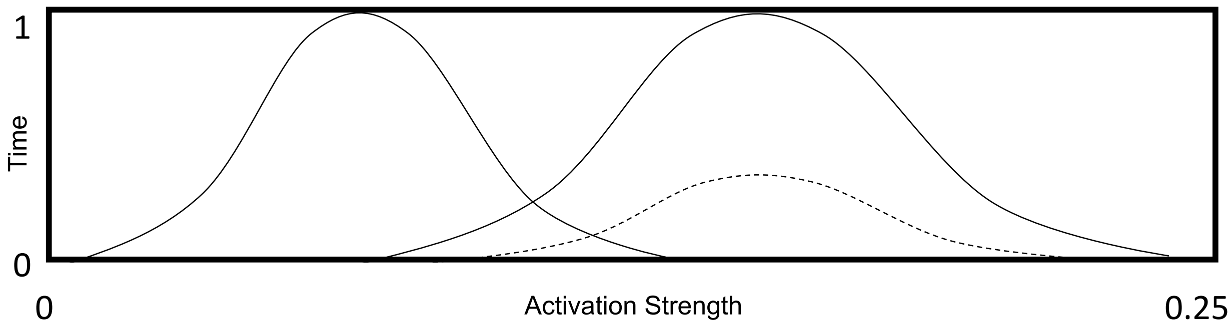


Figure 14 (d)

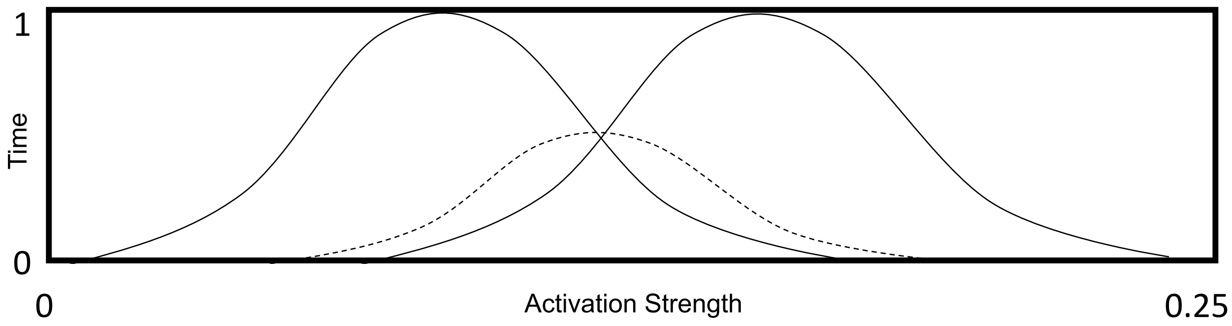
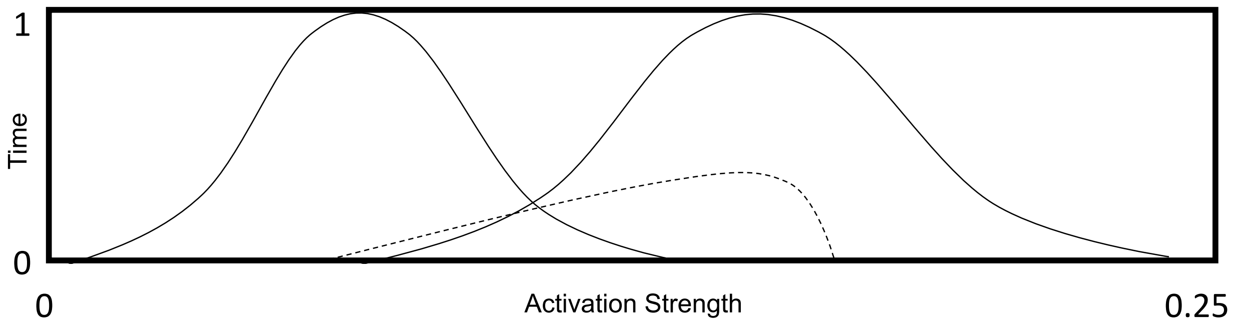


Figure 14 (e)



Form #	Syllable grouping	Prosodic constituents	Example
1	4-2	[NP]-[V_O]	[骨科醫生] [返工] [*gwat7fo1ji1sang1*] [*faan1gung1*] “The Orthopedist goes to work.”
2	3-3	[V-NP]-[V-NP]	[煲雞湯] [加花生] [*bo1gai1tong1*] [*gaa1faa1sang1*] “Add peanuts when cooking the broth.”
3	2-2-2	[NP]-[V_O]-[V_O]	[家姐] [開燈] [溫書] [*gaa1ze1*] [*hoi1dang1*] [*wan1syu1*] “Sister switches on the light to study.”
4	2-1-3	[NP]-[V]-[NP]	[叮噹] [聽] [收音機] [*ding1dong1*] [*teng1*] [*sau1jam1gei1*] “Doraemon listens to the radio”
5	2-4	[VAsp]-[NP]	[分開] [堆污糟衫] [*fan1hoi1*] [*deoi1wu1zou1saam1*] “Separate that heap of dirty clothes.”

N =noun; V = verb; V_O = Verb-object phrase NP = noun phrase; Asp = aspect marker

Table 1: Sample stimuli for the speech cycling paradigm

Reference Level: *3-3*	**Syllable 3**	**Syllable 4**	**Syllable 5**
2-4	Later	Earlier	No difference
2-2-2	Later	Earlier	Later
4-2	No difference	Earlier	Later
2-1-3	Later	No difference	No difference

Table 2: Predicted patterns for syllable alignment compared to the reference condition (3-3 grouping condition) for each syllable and prosodic form condition. Cells highlighted in green reflect that results matched predictions; those highlighted in red reflect results violated predictions (lighter red shading indicates no significant difference was found across conditions).