Accepted for publication in the *Journal of the Acoustical Society of America*, March 2025

The impact of clear speech modifications on perceived tempo of rate-matched English utterances

Leendert Plug,[1] Yue Zheng,[1] and Rachel Smith[2]

[1] *Linguistics and Phonetics, University of Leeds, Leeds, United Kingdom*

[2] *Laboratory of Phonetics, University of Glasgow, Glasgow, United Kingdom*

This study investigates the impact of clear speech modifications on perceived tempo, when listeners judge utterances which differ in speaking mode but are matched for articulation rate. Previous research motivates competing hypotheses: clear stimuli should sound faster if listeners' judgements are informed by differences in rate of spectral change; or slower if listeners are informed by knowledge of production patterns and differences in intelligibility; or the same in tempo if listeners weigh the available cues equally. This study addresses these hypotheses in three experiments using paired clear and normal productions of English sentences, with the clear productions manipulated to match their articulation rate to that of the normal productions. Experiment 1 assessed listeners' ability to separate the parameters of tempo and speaking mode when exposed to these pairs. Experiment 2 assessed the perceived direction of tempo difference in the same pairs. Experiment 3 expanded on Experiment 2 by adding two dimensions of variability: productions of different sentences were paired, and articulation rate was further manipulated to yield slow, mid and fast pairs. Across the experiments, the hypothesis that rate-matched clear stimuli should sound faster than normal ones finds the strongest support. Implications for our understanding of tempo perception are discussed.

I. INTRODUCTION

Research on the perception of speech tempo has identified multiple acoustic cues for tempo variation. Syllable rate seems the most robust predictor of perceived tempo (Grosjean and Lane, 1974; Grosjean, 1977; Pfitzinger, 1999; Gibbon *et al.*, 2015), but non-temporal cues are also present: increases in vowel space size and *f*0 and intensity spans and levels all raise perceived tempo (Feldstein and Bond, 1981; Kohler, 1986; Rietveld and Gussenhoven, 1987; Weirich and Simpson, 2014). This suggests that listeners are sensitive to the rate of spectral change in a signal: the higher this is within a given time window, the higher the perceived tempo (Weirich and Simpson, 2014).

Interestingly, variation in the same acoustic parameters is implicated in variation along the 'H&H continuum' (Lindblom, 1990; 1996) and shifts between 'normal' and 'clear' speaking modes—albeit in a complex way. According to Lindblom, speakers adjust their articulatory precision between loosely controlled 'hypo-articulation' and firmly controlled 'hyper-articulation' depending on the need for clarity given situational constraints. Research that has focused on how speakers adjust their articulations when asked to speak clearly has highlighted recurrent correlates of clear speech, including decreased articulation rate, increased *f*0 span and level, increased intensity level and greater dispersion of vowels in the F1–F2 space (Bradlow *et al.*, 1996; Krause and Braida, 2004; Hazan and Baker, 2011; Hazan *et al.*, 2018). Clear speech is also associated with decreased coarticulation, yielding more easily delimitable phones and more canonical articulations (Picheny *et al.*, 1986; Matthies *et al.*, 2001; Smiljanić and Bradlow, 2008; Searl and Evitts, 2013).

Clear speech would therefore appear to combine cues for low perceived tempo with cues for high perceived tempo: it is relatively slowly articulated, but with increases in *f*0, intensity and vowel dispersion relative to normal articulation, as well as other articulatory adjustments which increase the spectral change in the signal. On the face of it, this motivates the hypothesis that when normal and clear utterances are rate-matched, the clear utterances should sound faster.

We should note two things, however. First, several studies have pointed to a positive correlation between perceived tempo and cognitive load (Bosker *et al.*, 2017; Bosker and Reinisch, 2017). If anything, the acoustic characteristics of clear speech reduce cognitive load: in line with 'H&H' theory, they enhance intelligibility (Picheny *et al.*, 1985; Smiljanić and Bradlow, 1999; Krause and Braida, 2002; Smiljanic and Gilbert, 2017; Ferguson and Morgan, 2018). This enhancement might therefore bias listeners towards hearing rate-matched clear speech as relatively slow. Second, when listeners recognise speech as clear, they may be biased towards hearing it as relatively slow because it usually *is* articulated slowly in ordinary speech. Several studies appeal to this reasoning. Koreman (2006) reports that phone deletions have a consistently positive effect on perceived tempo, and Reinisch (2016) finds that a naturally fast utterance version, with several phone deletions, is perceived as faster than a linearly compressed version of the same utterance, without deletions. Both suggest that their listeners have drawn on their knowledge that phone deletions generally occur in relatively fast speech. Boltz (2011) analogously explains listeners' bias towards hearing high-pitched, loud and staccato music as fast.

In this study, we probe the impact of clear speech modifications on perceived tempo in three related experiments. Crucially, listeners compare rate-matched normal and clear utterances. Given the above, we can formulate several competing hypotheses. If listeners' judgements are strongly informed by differences in rate of spectral change, clear stimuli should sound faster (Hypothesis 1). If they are strongly informed by knowledge of production patterns and differences in intelligibility, clear stimuli should sound slower (Hypothesis 2). If listeners weigh the available cues equally, they may report hearing no difference (Hypothesis 3). Note that a 'no difference' response could also be taken to reflect listeners' ability to ignore clear speech cues and accurately recognise articulation rate equality. We deem it unlikely that this is generally the best interpretation, given the results of previous studies with rate-matched stimuli. Still, no study to date has assessed listeners' ability to separate rate differences from other clear speech adjustments *if asked*. We do this in Experiment 1.

II. EXPERIMENT 1

In most tempo perception experiments, listeners are asked to judge rate-matched stimuli on the single perceptual dimension of tempo. It is possible that in such designs, which we use in Experiments 2 and 3, listeners are biased to map any perceived acoustic variation to the one response dimension. In Experiment 1 we gauged the extent to which listeners can separate speech tempo from other aspects of speaking mode variation when multiple response dimensions are available to them.

A. Method

1. Participants

The online platform Prolific (www.prolific.co) was used to recruit 82 native speakers of British English aged 18 to 35 (44 female, 24 male, 14 non-specified). All were English monolinguals with no known language-related disorders or hearing impairment. All passed a short screening task in which they discriminated between sentence pairs with identical and non-identical members.

2. Stimuli

a. Corpus. We created stimuli using the materials in the *LUCID Corpus* (Hazan and Baker, 2011), accessed through *SpeechBox* (Bradlow, n.d.). These include a set of 144 sentences read by 40 Southern Standard British English speakers in normal and clear speaking modes. For the former, speakers read the sentences 'casually, as if talking to a friend'; for the latter, 'clearly as if talking to someone who is hearing impaired'.

b. Sentence and speaker selection. We selected ten sentences (Table 1 of Supplementary Materials) as produced by six speakers (four female, two male). Because our interest was in the rate of spectral change, we chose speakers who did not make large initiation effort and voice quality adjustments in speaking clearly, but whose clear productions still had distinct articulatory and prosodic characteristics. Each speaker produced all sentences.

c. Phonetic analysis. We segmented each sentence production (N=120) with *G2P* and *WebMAUS* (Kisler *et al.*, 2017), using the 'English (GB)' language model, producing a *Praat* TextGrid (Boersma & Weenink, 2017) with a surface phone-level segmentation. Boundary placements for vowels were checked and manually corrected where relevant. We added an editable *f*0 contour using *Mausmooth* (Cangemi, 2015) (step 0.05s, range 15–400Hz). We then extracted a number of acoustic parameters, largely following Hazan and Baker (2011): total duration and (canonical) syllable rate (cf. Plug *et al.*, 2021); Long-Term Average Spectrum (LTAS); F1 and F2 median and range across stressed monophthongal vowels; *f*0 median and standard deviation; and proportional vowel duration (%V). (We did not analyse intensity as this was normalised in the recordings.) For LTAS, we followed Hazan and Baker (2011). We fitted a linear mixed-effects model for each measure, using the *lme4* package (Bates *et al.*, 2015) in R (R Development Core Team, 2008), with Mode ('normal' *vs* 'clear') as fixed effect and random intercepts for Sentence and Speaker. This confirmed significant effects of Mode for duration, articulation rate, LTAS and *f*0 standard deviation. Means and standard deviations across speakers are in Table I; values per sentence are in Table 1 of Supplementary Materials. Across speakers and sentences, all clear productions are longer and more slowly articulated; most have a higher LTAS; and most have greater *f*0 dispersion.

TABLE I. Means (standard deviations in parentheses) for the principal distinguishing parameters between clear and normal sentence productions.

	Duration (sec)	Articulation rate (sylls/sec)	LTAS (dB)	*f0* standard deviation (ERB)
normal	1.54 (0.17)	4.71 (0.64)	40.1 (5.44)	1.09 (0.50)
clear	2.29 (0.31)	3.21 (0.47)	47.5 (6.58)	1.33 (0.54)

Moreover, we noted systematic articulatory differences between the sentence productions, consistent with descriptions of British English connected speech phonetics (Collins & Mees, 2013; Ogden, 2009). Normal sentence productions containing *the*, *at*, *from*, *of*, *was* and *a* consistently contain 'weak form' pronunciations—[ðə], [ət], [fɹəm], [əv], [wəz], [ə] respectively—while clear productions often contain 'strong forms'—[ði], [æt], [fɹɒm], [ɒv], [wɒz], [eɪ]. Several instances of phone deletion are apparent in the normal productions—/t/-deletion in *lost my* [lɒsmaɪ], /d/-deletion in *old lady* [ɵʊleɪdɪ], /ə/-deletion in *to the* [tðə] and /h/-deletion in *gave his* [ɡeɪvɪz]—which are absent in the clear productions. In total, 43 clear sentence productions (72%) differed from their corresponding normal production in at least one of these easily identifiable phone-level characteristics. In addition, in the normal sentence productions, [t], [k] and [ɡ] are mostly unreleased when followed by another consonant, as in *seat came* [sit̚kʰeɪm]; in the clear productions they are mostly released. Lack of /t/-release is often accompanied by place assimilation: e.g. *ate the* [eɪt̪̚ðə]; when /t/ is fully released, its closure is generally alveolar. One of the male speakers even produces ejective releases in clear sentence productions: e.g. *the sheep* [ðəʃip']. Vowel-initial words such as *old, ate* and *of* mostly start with a glottal closure in clear productions, as in *lady ate* [leɪdɪʔeɪt]; in normal ones, the hiatus is managed without glottalization, as in *lady ate* [leɪdɪjeɪt].

In terms of prosodic phrasing, each normal sentence production forms one Prosodic Phrase (PP), as Smiljanić and Bradlow (2008) call it, with an intonation contour appropriate for a neutral statement and no obvious breaks in articulation. In most clear productions, by contrast, either the first foot forms its own PP, or each foot does: e.g. *She's going* | *to sue the firm*, *The suit* | *was full* | *of holes*. Some of the clear speech features described above contribute to this phrasing: for example, /t/-release in *suit* and glottal closure starting *of* in *The suit* | *was full* | *of holes*. Seven clear sentence productions (12%) contain a perceived pause marking a PP boundary. Since phrase-internal pausing is highly relevant for speech tempo perception (e.g. Grosjean and Lane, 1974), we manually reduced these silences' durations. This did not change the observations that all clear productions are longer and more slowly articulated than normal productions, with more PPs.

d. Sentence manipulation and pairing. For each sentence, produced by each speaker, we created four sentence production pairs (Table II). Pair members were always by the same speaker.

TABLE II. Stimulus pair types for Experiment 1; '~' indicates that pairs were included in both possible orders.

Type	Pair members	Syllable rate	Speaking mode
SPEED	clear$_{\text{natural}}$ ~ clear$_{\text{compressed}}$	different	same
PRECISION	normal$_{\text{natural}}$ ~ clear$_{\text{compressed}}$	same	different
BOTH	normal$_{\text{natural}}$ ~ clear$_{\text{natural}}$	different	different
NEITHER	normal$_{\text{natural}}$ ~ normal$_{\text{natural}}$ clear$_{\text{natural}}$ ~ clear$_{\text{natural}}$	same	same

In SPEED pairs, the two productions differ in syllable rate, but not in speaking mode. This was achieved by pairing the (natural) clear production with a clear production whose duration was compressed to that of the speaker's corresponding normal production. In PRECISION pairs, the two productions differ in speaking mode, but not in syllable rate, achieved by pairing the (natural) normal production with the same compressed-to-normal clear production. In BOTH pairs, the two productions differ in both speaking mode and syllable rate, achieved by pairing the normal and clear productions without manipulation. In NEITHER pairs, the two productions are identical—that is, different in neither speaking mode nor syllable rate, achieved by pairing the (natural) normal production with itself, or the (natural) clear production with itself.

Compressions were done with PSOLA resynthesis in Praat (Boersma & Weenink, 2017). To avoid order effects, each pair was created in both possible orders, except the NEITHER pairs, which contained the identical token twice. Pair members were separated by a 1-second silence.

3. Task and procedure

a. Task and expected response. The participants' task was to decide for each sentence production pair which of the following descriptions best fitted their impression of the relationship between the pair members: a) Different in the speed of articulation; b) Different in the precision of articulation; c) Different in both; d) Different in neither. These response options map onto the four types of production pair straightforwardly: in SPEED pairs, the two productions differ in articulation rate, so 'a' (henceforth a 'speed' response) is signal-consistent; in PRECISION pairs, the two productions differ in speaking mode, so 'b' ('precision') is signal-consistent; in BOTH pairs, the two productions differ on both parameters, so 'c' ('both') is signal-consistent; in NEITHER pairs, the two productions do not differ, so 'd' ('neither') is signal-consistent. We preferred the term 'precision' over 'clarity' as we wanted to focus listeners' attention on the speaker's articulation, not the production's intelligibility.

Our particular interest was in how participants responded to PRECISION pairs. Substantial identification of PRECISION pairs as differing in 'speed' or 'both' would confirm that speaking mode differences trigger tempo percepts in the absence of articulation rate differences, in line with our Hypotheses 1 and 2. The inclusion of SPEED pairs allows us to assess whether the opposite pattern also occurs; based on previous experimental findings, we expected that listeners would recognise linear rate manipulations as temporal differences alone.

b. Procedure. To keep the number of trials per participant manageable, we created four lists of 120 trials, counterbalancing for sentence and within-pair order. Two lists contained all the sentence production pairs (spoken by all six speakers) for five out of the ten sentences, while the other two lists contained all the pairs (spoken by all six speakers) for the other five sentences. For each pair of types SPEED, PRECISION or BOTH, one list contained the pair in one order—e.g. with the manipulated production in first position—and another list contained it in the other order—e.g. with the manipulated production in second position. For each NEITHER pair, one list contained two normal sentence productions and one list contained two clear sentence productions. Participants were randomly allocated to the lists (n=20 or n=21 for all lists).

We used the *Gorilla Experiment Builder* (www.gorilla.sc) to build and run the experiment online (Anwyl-Irvine *et al.*, 2020). At the start of the online session, participants were given information about the experiment and were asked to confirm that they met the participant recruitment criteria and agreed to take part in the experiment.

Participants were instructed to listen to the sentence production pairs and judge whether the pair members sounded the same, different in articulation speed, different in articulation precision, or different in both speed and precision. In each trial (N=120), the pair played twice with a 2-second silence, while participants saw a screen displaying the sentence, an audio icon, and a reminder of the core instruction. The screen then changed to display the four response options 'speed', 'precision', ‘both’, and ‘neither’. The next trial began automatically once the participant had submitted a response. There was a 0.5-second silence between trials. The order of trials was randomised for each participant. Participants were allowed to take a short break after completing 40 and 80 trials. The experiment took approximately 30 minutes.

B. Results

Figure 1 shows the response proportions by stimulus type. Listeners are very good at identifying NEITHER pairs (93% ‘neither’). For SPEED pairs, ‘speed’ is also the majority response (64% ‘speed’). For PRECISION pairs, ‘precision’ is the majority response (50% ‘precision’), but this majority is smaller than that for SPEED pairs. Participants are least accurate at identifying BOTH pairs (40%): a majority of responses suggests participants were hearing difference in one parameter only (33% ‘precision’, 22% ‘speed’).

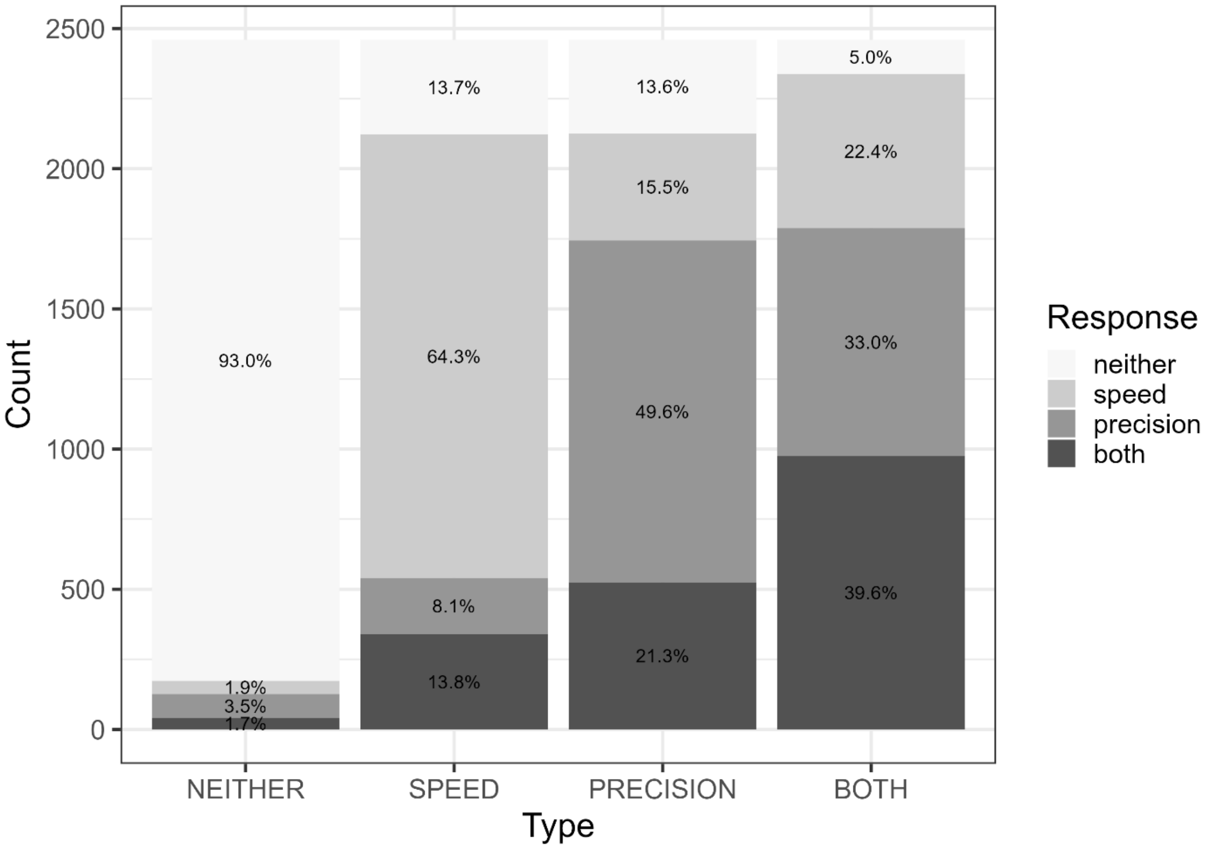

FIG. 1. Numbers of ‘neither’, ‘speed’, ‘precision’ and ‘both’ responses to NEITHER, SPEED, PRECISION, and BOTH stimuli in Experiment 1.

For statistical analysis we decomposed the responses into ‘speed_accuracy’ (1 for ‘speed’/‘both’ responses to SPEED/BOTH pairs and ‘precision’/‘neither’ responses to PRECISION/NEITHER pairs; 0 otherwise), and ‘precision_accuracy’ (1 for ‘precision’/‘both’ responses to PRECISION/ BOTH pairs and ‘speed’/‘neither’ responses to SPEED/NEITHER pairs; 0 otherwise). We modelled each using mixed-effects logistic regression, using the *glmer*() function in *lme4*, with Participant and Sentence as random variables, and Type (levels: SPEED, PRECISION, BOTH, NEITHER) as a fixed predictor.

Type significantly predicted accuracy on the ‘speed’ dimension. Speed_accuracy was significantly lower ($\beta = -0.76221$, $z = -11.65$, $p < 0.0001$) on PRECISION trials than on SPEED trials. This reflects that 16% of PRECISION stimuli attracted ‘speed’ responses and 21% attracted ‘both’ responses. Accuracy was also significantly lower (β = -0.81482, z = -12.48, p < 0.0001) on BOTH than SPEED trials, despite an observable rate difference being present in both. 33% of BOTH trials received ‘precision’ responses, suggesting that simultaneous variation in speaking mode can cause listeners to conflate ‘speed’ and ‘precision’.

Type also significantly predicted accuracy on the ‘precision’ dimension, but the pattern of responses was different. Precision_accuracy was higher (β = 0.40027, z = 5.92, p < 0.0001) on SPEED trials than on PRECISION trials. This reflects that participants were less likely to hear articulation rate differences as involving ‘precision,’ compared with hearing speaking mode differences as involving ‘speed’. Precision_accuracy on BOTH trials did not significantly differ from that on PRECISION trials (β = 0.09379, z = 1.44, n.s.), suggesting that simultaneous variation in speed did not affect participants’ observation of speaking mode variation.

C. Discussion

In relation to our central hypotheses, the Experiment 1 results confirm that when exposed to rate-matched stimuli that differ in speaking mode, listeners report hearing tempo variation in a substantial minority of instances (37%), even if they are encouraged to separate ‘speed’ and ‘precision’ judgements. We believe that this response pattern motivates further investigation of the direction of perceived difference, to establish whether Hypothesis 1 or 2 is better supported. We also note that listeners accurately reported hearing no tempo difference and only a difference in ‘precision’ in 50% of instances. In relation to Hypothesis 3, this suggests that listeners’ ability to identify that sentences are rate-matched when differing in speaking mode is limited. More broadly, we take the results of Experiment 1 to highlight the close and complex relationship between perceived tempo and perceived articulatory ‘precision’. We will return to this relationship, and its implications for experimental design, in the General Discussion.

III. EXPERIMENT 2

In Experiment 2 participants heard the same pairs as in Experiment 1 and judged how, as well as whether, the productions differed in tempo. In line with previous experimental studies (Feldstein and Bond, 1981; Kohler, 1986; Rietveld and Gussenhoven, 1987; Weirich and Simpson, 2014; Plug and Smith, 2021), participants were not alerted to the variation in speaking mode. This allowed us to address our hypotheses 1 and 2 regarding the direction of any effect of clear speech acoustics on perceived tempo when articulation rate is controlled.

D. Method

1. Participants

Participants were recruited from student cohorts at the University of Leeds. We tested 41 native speakers of British English aged 18 to 35 (27 female, 7 male, 7 non-binary or unreported) with no known language-related disorders or hearing impairment. None had participated in Experiment 1. All passed the same short screening task as used in Experiment 1. Most completed the experiment in supervised lab sessions and received a small payment; some completed it remotely.

2. Stimuli

The stimuli and experimental lists were the same as in Experiment 1 (n=9, 10 or 11 participants per list).

3. Task and procedure

a. Task and expected responses. The participants' task was to decide for each stimulus pair which of the following descriptions best fitted their impression of the relationship between the pair members: a) The first production is faster; b) The second production is faster; c) Neither is faster.

Given the results of Experiment 1 and previous studies, we expected participants to identify SPEED pair members with higher articulation rates as faster, and to identify the absence of any difference in NEITHER pairs with high degrees of accuracy. We expected participants to report hearing a tempo difference in at least a substantial minority of PRECISION trials (cf. 37% in Experiment 1). We also expected participants to report hearing a tempo difference in a similar proportion of BOTH trials as that observed in Experiment 1 (62%).

b. Procedure. The procedure was identical to Experiment 1 except for the screen display. Participants were instructed that in each of the pairs they were going to hear, the members might sound the same, or one member might sound faster. In each trial (N=120), the screen display showed the sentence text, an audio icon, and the question 'Which one of the two productions is faster?'; it then changed to show the response options 'Production 1', 'Production 2' and 'Neither'. The experiment took approximately 30 minutes.

E. Results

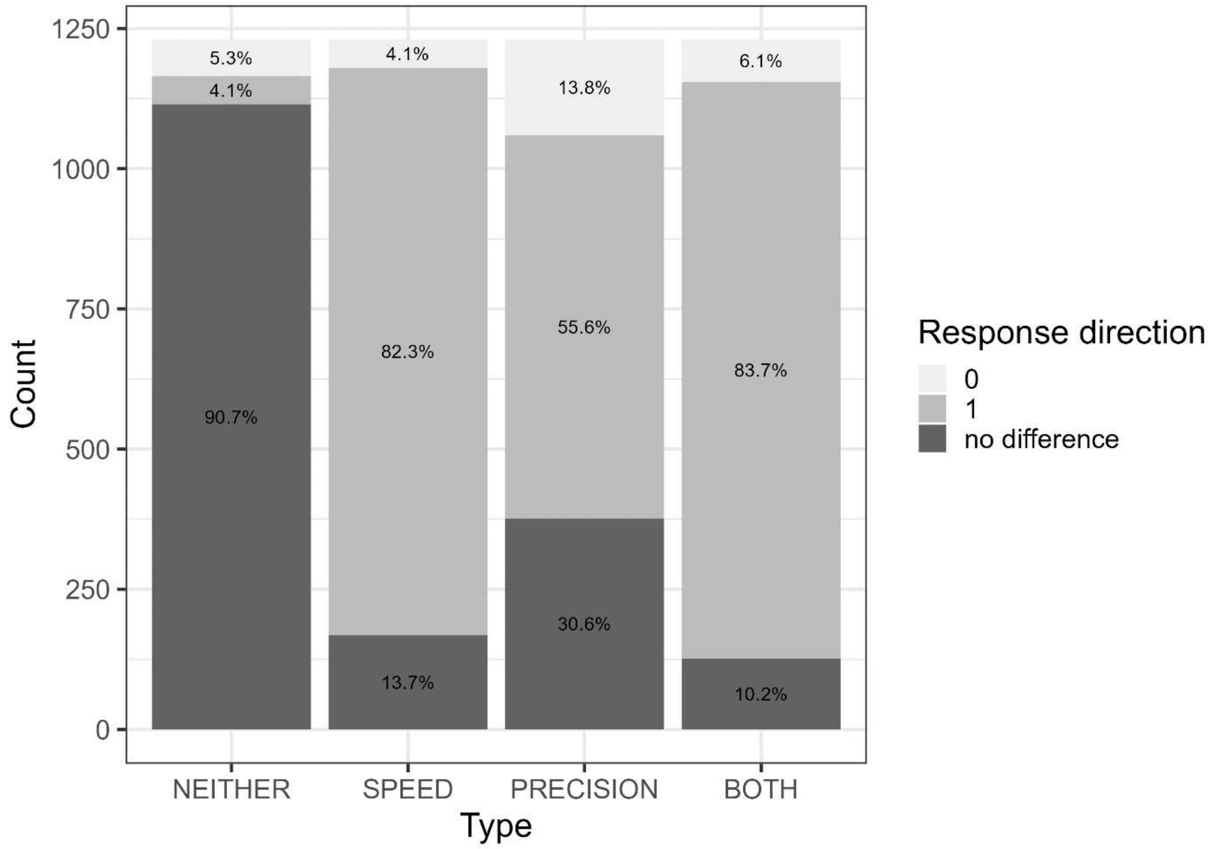

FIG. 2. Responses summary for Experiment 2. Responses were coded 1 where a 'faster' response was given to the clear pair member (PRECISION trials), or the member with the higher

rate (SPEED and BOTH trials), or the first member (NEITHER trials). Responses in the opposite direction were coded 0.

Figure 2 shows that 69.4% of responses to PRECISION pairs report a perceived tempo difference: 55.6% suggesting that the clear pair member is faster (supporting Hypothesis 1), and 13.8% suggesting the opposite (supporting Hypothesis 2). 30.6% report 'neither' (i.e., no perceived difference in tempo), which offers some support for Hypothesis 3 (though we cannot rule out that successful perceptual separation of tempo and mode variation underlies some of these responses). For SPEED and BOTH pairs, 86.4% and 89.8% of responses identify a tempo difference; almost all in the expected direction of pair members with higher articulation rates being faster. As in Experiment 1, listeners identified NEITHER pairs with accuracy of above 90%.

While these results offer most support for Hypothesis 1, we should highlight that listeners were less likely to perceive the clear stimulus in PRECISION trials as faster in its pair, than to perceive the higher-rate stimulus in SPEED trials as faster. This is confirmed by a mixed-effects logistic regression analysis. Excluding 'neither' (i.e. no perceived difference) responses, we coded 'higher rate faster' responses to SPEED and BOTH pairs '1' and responses in the opposite direction '0'. For PRECISION pairs, we coded 'clear faster' responses '1' and responses in the opposite direction '0'. A model with participant and sentence as random effects and Type as a factor confirms that listeners made fewer responses in the '1' direction in PRECISION than SPEED trials ($\beta = -1.7586$, $z = -10.073$, $p < 0.0001$). They also made very slightly but significantly fewer responses in the '1' direction in BOTH trials compared to SPEED trials ($\beta = -0.3894$, $z = -2.045$, $p = 0.04$); this suggests that the co-presence of clear speech characteristics along with lower articulation rate slightly reduced the likelihood of participants reporting slower tempo for the clear member.

We investigated whether the acoustic variables that differentiate clear and normal sentence productions might explain why some clear members of PRECISION pairs were heard as slower than their normal counterparts. For each PRECISION pair, we calculated the difference between the compressed-clear and normal production on the four significant variables from the acoustic analysis (see II.A.2.c): *f*0 standard deviation, LTAS, duration and articulation rate of the original uncompressed clear token. The derived difference measures were centred and scaled. We again coded "clear faster" responses to PRECISION stimuli '1', and "clear slower" responses '0'. In a logistic regression with random effects for participant and sentence, the difference in *f*0 standard deviation positively predicted likelihood of a "clear faster" response (β = 0.26298, z = 2.582, p < 0.01) as did the duration difference (β = 0.36184, z = 2.254, p < 0.025). This means that clear tokens that were heard as slower tended to have undergone relatively little compression to achieve rate-matching, and contained notably little *f*0 variation. The other predictors were not significant.

F. Discussion

Experiment 2 sought to establish the direction of the tempo perception variation that listeners experience when listening to sentence productions whose rate is controlled, but which vary in speaking mode. Results show that with rate controlled and listeners given only a single response parameter, clear speaking mode makes speech sound relatively fast in over 60% of trials. In a substantial minority of trials, it yields a percept of no tempo difference, and in a small minority of trials, even a slower percept. Hypothesis 1 is therefore most clearly supported; however, Hypotheses 2 and 3 cannot be discarded, and it is possible that listeners are sensitive to fine phonetic detail that was not controlled in our stimuli. We will return to this point in the General Discussion. Moreover, higher articulation rate, with speaking mode controlled, yields a percept of faster tempo more reliably than clear speech characteristics do.

IV. EXPERIMENT 3

Experiment 3 sought to assess the robustness of the main findings of Experiment 2 and explore the premise of our Hypothesis 2—'listeners' judgements are strongly informed by knowledge of production patterns'—further. We recruited additional participants and used a stimulus set which differed from Experiment 2 in two ways. First, we used stimulus pairs consisting of normal and clear productions of two different sentences, as opposed to two productions of the same sentence. Second, we manipulated the articulation rates of all stimuli to yield slow, mid-tempo and fast sentence production pairs. Because of the general associations between slow speech and hyper-articulation on the one hand, and fast speech and hypo-articulation on the other, Hypothesis 2 can be taken to predict that effects of speaking mode on perceived tempo will be more salient at fast and slow tempi compared with mid-range tempo (see also Koreman, 2006).

G. Method

4. Participants

We again used Prolific (www.prolific.co) to recruit 99 native speakers of British English aged 18 to 35 (60 female, 32 male, 7 non-specified). None had participated in Experiments 1 or 2. Participation criteria and the screening task were identical to Experiments 1 and 2.

5. Stimuli

a. Corpus, sentence and speaker selection. We again used the *LUCID Corpus* (Hazan and Baker, 2011). We selected just one of the six speakers used previously. The *LUCID Corpus* contains speakers' productions of sentence pairings matched for general phonological make-up. We selected nine pairings whose sentences were produced fluently with the articulatory and prosodic characteristics of normal *vs* clear speech described for Experiment 1.

b. Sentence manipulation and pairing. We created multiple sentence production pairs for each sentence pairing in Table III. First, we generated three versions of each sentence production: one slow, one mid-tempo and one fast. To create the mid-tempo versions, we calculated the mean duration across the unmanipulated normal productions of the two sentences in each sentence pairing; we then set the durations of both sentences' normal and both sentences' clear productions to this mean value. This ensured that speaking mode was not confounded with syllable rate. To create the slow versions, we multiplied the mid-tempo target durations by 1.3; for the fast versions by 0.7. These sizeable changes, resulting in mean articulation rates of 5.52 (fast), 3.87 (mid), and 2.97 (slow) syllables per second, allowed us to assess the perceptual impact of speaking mode variation at fairly extreme tempi. Manipulations were done through PSOLA resynthesis in *Praat* (Boersma & Weenink, 2017).

TABLE III. Sentence pairs used in Experiment 3 and their production duration and syllable rate under each speech mode.

Pair	Sentence	Duration (sec)		Articulation rate (sylls/sec)	
		normal	clear	normal	clear
1	*My brother Paul ran towards the beach.*	2.19	2.79	4.11	3.23
	The bouncy ball rolled towards the sea.	2.07	2.61	4.35	3.45
2	*The pear belongs to the teacher.*	1.84	2.03	4.35	3.94
	The bear belongs to the children.	1.67	1.96	4.79	4.08
3	*The peas were shelled in a bowl.*	1.57	2.09	4.46	3.35
	The bees were kept on a farm.	1.70	1.95	4.12	3.59
4	*Daisy the sheep was grazing.*	1.69	1.98	4.14	3.54

	Birds in the sea are noisy.	1.66	2.05	4.22	3.41
5	*My mother paints foreign shells.*	1.82	2.42	3.85	2.89
	My father clones human cells.	2.06	2.50	3.40	2.80
6	*After work he went to the shop.*	1.73	2.27	4.62	3.52
	After school she knitted a sock.	1.84	2.36	4.35	3.39
7	*The sleeping pill was very effective.*	1.90	2.40	5.26	4.17
	My heating bill is very expensive.	2.02	2.53	4.95	3.95
8	*Jonathan gave his wife a bush.*	1.85	2.23	4.32	3.59
	Everyone gave the boat a push.	1.66	2.16	4.82	3.70
9	*The sheep were grey and old.*	1.54	1.90	3.90	3.16
	The sea was blue and calm.	1.50	1.89	4.00	3.17
Mean (sd)		1.80 (0.20)	2.23 (0.27)	4.33 (0.44)	3.50 (0.38)

We then created sentence production pairs of three types (Table IV). In PRECISION pairs, the two sentence productions differed in speaking mode only: we paired the normal production of one sentence (e.g. *The bear belongs to the children*) with the clear production of its matched sentence (e.g. *The pear belongs to the teacher*). We did this with the sentences in both possible orders, and with clear and normal productions in both possible orders. This creates 3 (Tempo: *slow, mid, fast*) x 2 (Position of clear stimulus: *first, second*) x 2 (Sentence order) = 12 production pairs for each sentence pairing.

TABLE IV. Sentence production pairing for Experiment 3.

Type	Pair members	Syllable rate	Speaking mode
PRECISION	normal ~ clear	same	different
SPEED	normal ~ normal clear ~ clear	different	same
NEITHER	normal ~ normal $\text{clear}_{\text{natural}}$ ~ $\text{clear}_{\text{natural}}$	same	same

In SPEED pairs, the two sentence productions differed in articulation rate only: we paired the normal or clear production of one sentence with the same-mode production of the matched sentence, with the articulation rates such that the two productions were either slow and mid (henceforth Tempo *slow*), or fast and mid (Tempo *fast*). We did this in both possible articulation rate orders. For these pairs, to keep experiment size manageable, we did not manipulate sentence order independently of articulation rate order: e.g. *The bear belongs to the children* was always the higher-rate member and *The pear belongs to the teacher* was always the lower-rate member. This creates 2 (Tempo: *slow, fast*) x 2 (Mode: *clear, normal*) × 2 (Position of higher-rate stimulus: *first, second*) = 8 production pairs for each sentence pairing.

In NEITHER pairs, the two sentence productions differed in neither speaking mode nor articulation rate. That is, we paired either two normal sentence productions with the same articulation rates or two clear productions with the same articulation rates. Again we did this with the sentences in both possible orders and at three general tempi. This again creates 3 (Tempo: *slow, mid, fast*) × 2 (Mode: *clear, normal*) × 2 (Sentence order) = 12 production pairs for each sentence pairing.

Sentence productions were separated by a 1-second silence within pairs.

6. Task and procedure

a. Task and expected responses. The participants' task was to decide for each sentence production pair which of the pair members was faster, and estimate the extent of difference. Our main interest was in PRECISION pairs. Based on Experiment 2, we expected that in PRECISION trials, the clear productions would predominantly be heard as faster than the normal ones. As indicated above, if listeners generally associate slow speech with hyper-articulation and fast speech with hypo-articulation, then hypo-articulated (normal) productions should be noticeably unusual at slow tempo, as should hyper-articulated (clear) productions at fast tempo. Therefore, our Hypothesis 2 can be taken to predict a stronger perceived tempo difference at fast and slow tempi than at mid-range tempo. SPEED and NEITHER pairs were included as controls: we expected listeners to be very consistent in identifying the articulation rate differences, or lack of differences, respectively.

b. Procedure. To keep the experiment length manageable, the 288 experimental items were split into three lists, counterbalancing for sentence pairs and tempo. Each list had 36 PRECISION items (12 at each of the three tempi), 24 SPEED items, and 36 NEITHER items (total = 96). Participants were assigned randomly to one of the three lists. After the screening described in Experiment 1, they were instructed to listen to pairs of sentences, decide which pair member sounded faster and estimate how much faster it was. Each pair played twice with a 2-second between-pair silence. The text of both sentences remained visible on screen throughout each trial. Participants registered their decisions using a slider whose midpoint was labelled "Same", with "Sentence 1 much faster" and "Sentence 2 much faster" at the left and right endpoints. After responding, they clicked a button to advance to the next trial. The main task started with 4 practice trials with 4 sentence pairs that were different from the experimental items. Trial order was randomized for each participant. The experiment took approximately 30 minutes.

H. Results

Slider responses were on a scale from -100 (“Sentence 1 much faster”) to +100 (“Sentence 2 much faster”). To analyse SPEED and PRECISION trials, we transformed responses so that a positive value always corresponded to a ‘faster’ judgement for the pair member that had higher articulation rate (on SPEED trials) or clear speaking mode (on PRECISION trials). For NEITHER trials, the untransformed responses were analysed.

Due to differences between the conditions in available variables or variable levels, we analysed each condition separately by fitting linear mixed-effects models, again using the *lme4* package in *R*. Fixed predictors were Tempo, Mode, and Position (of the higher-rate or clear stimulus within the pair). For each condition, we fitted a saturated model including all main and interaction effects allowed by the design, and the maximal random effects structure that yielded model convergence. We then applied backwards elimination of non-significant predictors using the *lmerTest* package (Kuznetsova *et al.*, 2017). Interaction plots were generated using the *emmeans* package (Lenth, 2024).

We first checked that participants responded in expected ways to NEITHER and SPEED pairs, i.e. detected no tempo differences in the former and robust tempo differences in the latter. Table V shows the optimal model for NEITHER pairs; Figure 3 plots model estimates. For these pairs, we expected a mean response of zero (“same”) and no effects of Mode or Tempo. For *mid* stimuli in both *normal* and *clear* speaking mode, and for *fast* stimuli in *normal* mode, listeners’ responses were in line with this expectation. For *slow* stimuli, however, responses were slightly but significantly skewed to the first pair member being heard as faster than the second. And for *fast* stimuli, in *clear* mode only, the second pair member was judged as slightly, but significantly, faster.

TABLE V. Summary of optimal model for NEITHER pairs in Experiment 3. Reference level of Mode is *normal*; reference level of Tempo is *mid*.

	Estimate	SE	df	*t*	*p*
(Intercept)	0.510	0.753	758.85	0.678	0.50
Mode *clear*	0.990	1.045	3362.02	0.948	0.34
Tempo *slow*	-3.606	1.050	2197.10	-3.435	<0.001
Tempo *fast*	0.027	1.093	484.89	0.025	0.98
Mode *clear* : Tempo *slow*	0.305	1.477	3362.02	0.206	0.84
Mode *clear* : Tempo *fast*	4.433	1.477	3362.02	3	<0.005

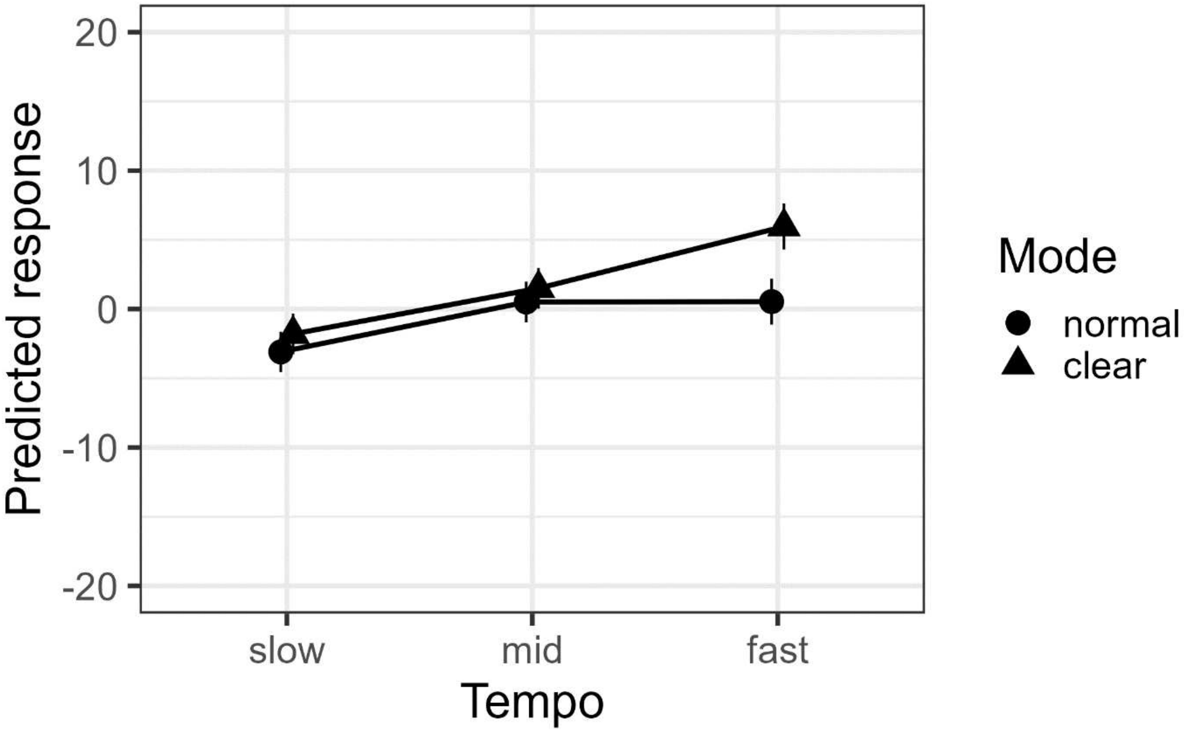

FIG. 3. Model-estimated responses to NEITHER pairs in Experiment 3, showing interaction of Tempo and Mode.

Table VI shows that for SPEED trials, the optimal model contained main effects of Tempo, Mode and Position of the higher-rate stimulus, interactions of Tempo with Position and Mode with Position, along with random slopes for Tempo by Participant and Item. Figure 4, top panel, shows that the pair member with the higher articulation rate was indeed reliably heard as faster. In *slow* pairs the perceived tempo difference was relatively small, averaging just under 30% of the 100-point scale, whereas in *fast* pairs it was significantly larger, averaging over 40% of the scale. This might be attributed to the nature of our rate manipulations: the duration ratio between the members of the pair was smaller in *slow* pairs (1.3:1 = 1.3) than in *fast* pairs (1:0.7 = 1.42). The interactions indicate that the order in which pair members were presented modulated the response patterns. The interaction of Position with Tempo (Figure 4, top) reflects that the difference between *fast* and *slow* pairs was greater when the higher-rate stimulus was in second position. An unexpected interaction of Position with Mode (Figure 4, bottom) reflects that when the higher-rate stimulus was in first position, responses to *clear* and *normal* pairs did not differ, whereas when the higher-rate stimulus was in second position, *clear* speaking mode slightly enhanced the perceived tempo difference between pair members, relative to *normal* mode.

TABLE VI. Summary of optimal model for SPEED pairs in Experiment 3. Reference levels are Mode *normal*, Tempo *slow*, and Position (of higher-rate stimulus) *first*.

	Estimate	SE	df	*t*	*p*
(Intercept)	27.59	2.079	24.48	13.271	<0.0001
Tempo *fast*	14.66	3.333	9.49	4.398	0.002
Mode *clear*	1.23	1.181	971.81	1.042	0.30
Position *second*	-4.77	1.446	1521.79	-3.300	<0.001
Tempo *fast* : Position *second*	6.05	1.644	2163.93	3.679	<0.001
Mode *clear* : Position *second*	4.16	1.704	333.06	2.441	<0.02

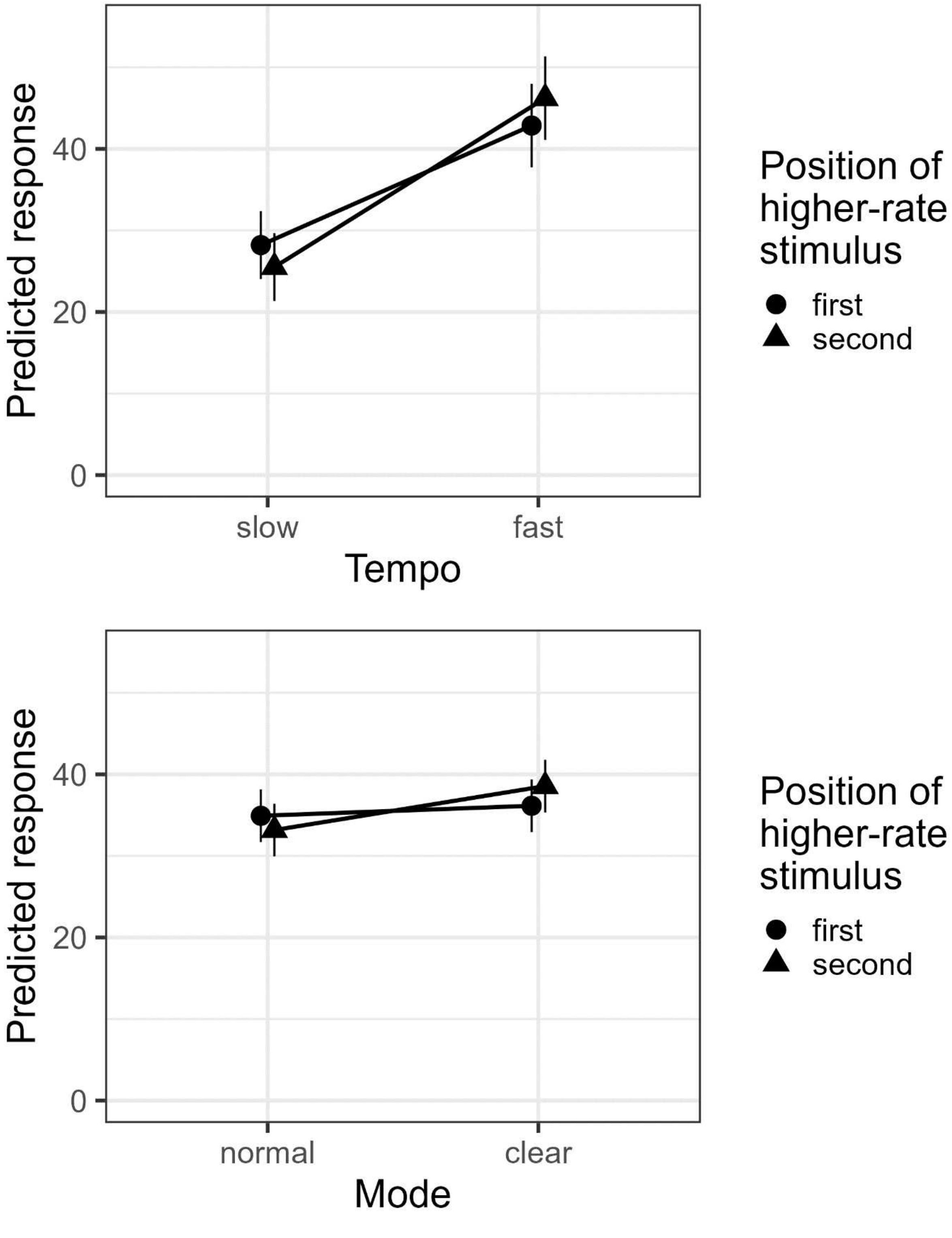

FIG. 4. Model-estimated responses to SPEED pairs in Experiment 3. Top: interaction of Tempo by Position (of higher-rate stimulus). Bottom: interaction of Mode by Position.

Turning to our main questions—whether participants detected tempo differences in PRECISION pairs and whether this was affected by global tempo—Table VII shows that the optimal model for PRECISION trials included main effects of Tempo, Position (of the clear pair member), and their interaction, as well as random intercepts for Participants and Items. Recall that these pairs consisted of a *clear* and a *normal* member, matched in rate. Figure 5 shows that, as predicted, the *clear* pair member was generally heard as faster than the *normal* member, at all tempi. Also as predicted, the *clear* pair member sounded faster than the *normal* to a greater extent for *slow* and *fast* compared to *mid* pairs. An unexpected interaction of Tempo and Position reflects one exception to this pattern: at *slow* tempo, when the *clear* member was heard second, the perceived tempo difference between *clear* and *normal* was substantially reduced.

TABLE VII. Summary of optimal model for PRECISION pairs in Experiment 3. Reference levels are Tempo *mid*, and Position (of clear stimulus) *first*.

	Estimate	SE	df	*t*	*p*
(Intercept)	14.99	1.5	30.96	9.989	<0.0001
Tempo *slow*	4.01	1.219	3455.04	3.291	0.001
Tempo *fast*	5.15	1.219	3455.04	4.225	<0.0001
Position *second*	1.29	1.217	3453.01	1.060	0.29
Tempo *slow* : Position *second*	-9.69	1.721	3453.02	-5.631	<0.0001
Tempo *fast* : Position *second*	2.40	1.721	3453.01	1.394	0.16

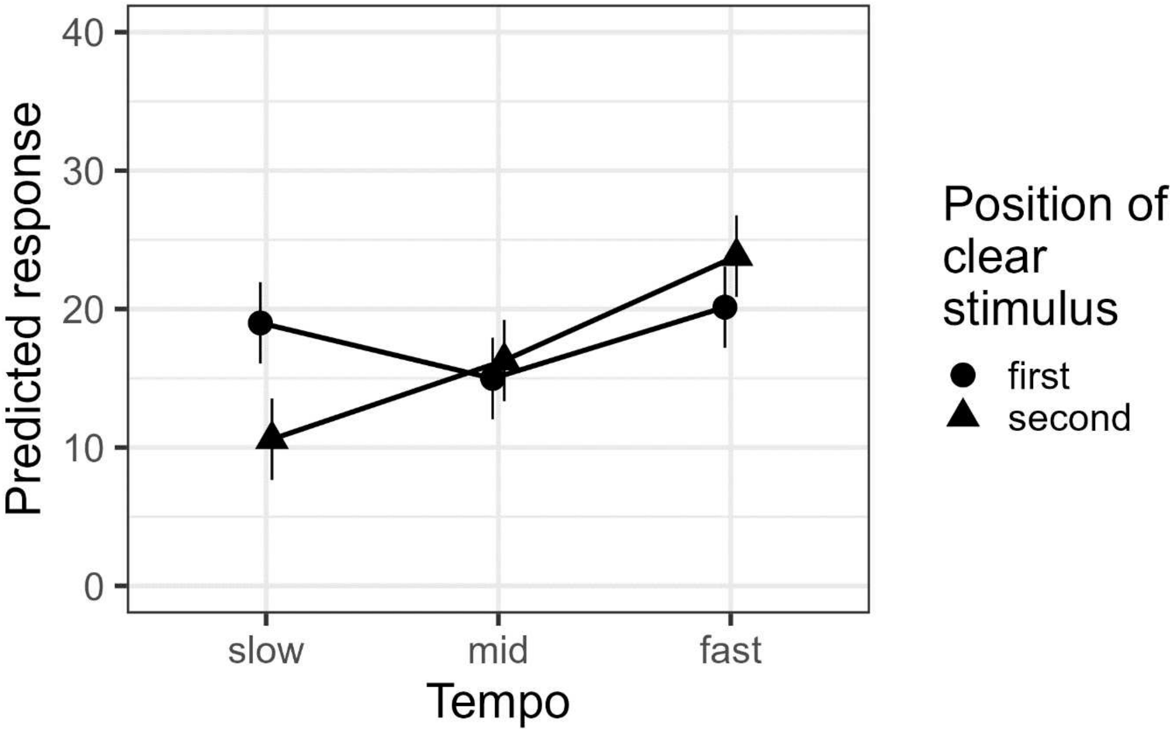

FIG. 5. Model-estimated responses to PRECISION pairs in Experiment 3, showing interaction of Tempo by Position (of clear stimulus).

I. Discussion

Experiment 3 set out to replicate and extend Experiment 2's finding that clear speech mostly sounds faster than rate-matched normal speech, testing across a range of tempi and with a more demanding task involving non-identical sentences. Results showed that listeners mostly judged clear sentence productions as sounding faster than rate-matched normal ones. Clear speech sounded faster across all tempi tested, but the effect was modulated by general tempo range: larger differences were heard at fast and slow tempi compared to mid. This provides support for our Hypothesis 2: when speech is notably slow, listeners strongly expect hyper-articulation, and when speech is notably fast, listeners strongly expect hypo-articulation; so in both cases they respond strongly to hearing the opposite. The unexpected exception to this general pattern is the interaction of speaking mode with the position of the clear sentence production: at slow tempo when the clear pair member is heard second, listeners perceive much less difference between clear and normal members than when the order is reversed. We can only speculate that this reflects a recency effect, in that when the last sentence heard has the features of stereotypically slow speech (low articulation rate, clear speaking mode) listeners are less likely to report it as faster than the first pair member. An expectation of final lengthening may also play a role here (cf. White *et al.*, 2012).

V. GENERAL DISCUSSION

We have reported on three experiments which probe the impact of clear speech modifications on perceived tempo when stimuli are matched for articulation rate. Experiment 1 tested listeners' ability to separate the parameters of 'speed' (tempo) and 'precision' (speaking mode). It revealed that they differentiate the parameters reasonably well when prompted to do so; however, they did report hearing differences in mode as involving tempo more frequently than *vice versa*. Experiment 2 established that when listeners are prompted to assess tempo only, they interpret speaking mode variation as tempo variation in a majority of trials, and mostly—although not exclusively—in the direction of hearing clear sentence productions as faster than rate-matched normal productions. Experiment 3 replicated this finding with sentence pairs that differed in linguistic content, and further showed that the effect was particularly strong in markedly high and low tempo ranges.

We pointed out at the outset that speaking mode variation presents an interesting puzzle with regard to perceived tempo. Intentionally clear speech is typically associated with a low speaking rate and hyper-articulation. In the stimuli we used, clear speech had significantly higher LTAS and *f*0 dispersion, fewer weak forms and deletions, more prosodic phrases, stronger plosive releases, more glottalized vowel onsets, and less assimilation compared with normal speech. These characteristics afford the speech high intelligibility and involve a degree of 'signal redundancy' (Aylett and Turk, 2004; Tang and Shaw, 2021). We could therefore quite reasonably formulate multiple, seemingly competing hypotheses as to how rate-matched normal and clear sentence productions would be perceived in relation to speech tempo. We explored three: clear sentence productions sound faster because of their higher rates of spectral change (Hypothesis 1); clear sentence productions sound slower because of listeners' knowledge of production patterns and because they are, if anything, more intelligible (Hypothesis 2); and clear sentence productions sound equal in tempo because listeners weigh the available cues equally (Hypothesis 3). We also examined the possibility that listeners' 'no tempo difference' responses might be accurate identifications of rate equality.

Across Experiments 2 and 3, we found clear support for Hypothesis 1: in both experiments, listeners reported hearing the clear sentence production as faster than its normal counterpart in a majority of trials. Our findings are therefore largely consistent with Weirich and Simpson's (2014) reasoning that an increase in the rate of spectral change in a signal makes it sound faster if its overall duration remains constant.

We found support for Hypothesis 2 too. First, in small minorities of trials, listeners reported hearing the clear sentence production as slower than the normal one. It is possible, then, that listeners' knowledge of typical production patterns and a speech signal's intelligibility constrain tempo judgements—but if this is the case, they only occasionally outweigh the relevance of factors related to the rate of spectral change. It is also possible that the responses to these trials were due to specific phonetic characteristics of the stimuli. Our phonetic analysis suggested that the clear members in the relevant pairs had narrow f0 ranges and had undergone relatively little compression: that is, they derived from relatively fast and monotonous clear sentence productions. In this regard, our findings add to those of previous tempo perception experiments in which listeners draw on available dimensions of variation in ways that are difficult to explain (Feldstein and Bond, 1981; Vitela *et al.,* 2013; Plug *et al.*, 2022).

Clearer support for the relevance of listeners' knowledge of production patterns comes from the finding in Experiment 3 that listeners report hearing more difference between rate-matched clear and normal sentence productions at low and high tempi compared with mid-tempo trials. This suggests that while listeners' knowledge of typical production patterns is generally outweighed by factors related to the rate of spectral change, knowledge that clear speech is typically slow does constrain the effect of the latter: when differences in the rate of spectral change map to unusual production patterns, their salience is increased. This is in line with the reasoning of Koreman (2006) and Reinisch (2016). To tease apart the possible effects of listener knowledge and intelligibility, we would need to develop designs in which intelligibility is manipulated more systematically, for example using noise-masked stimuli.

In relation to Hypothesis 3, we should highlight that the extent to which listeners reported hearing speaking mode variation as tempo variation differed considerably across our three experiments. In Experiments 2 and 3, listeners were only tasked to judge tempo, and 'difference' responses clearly outnumbered 'no difference' ones. In Experiment 1, listeners were prompted to separate 'speed' and 'precision', and 63% of responses in the crucial trials were 'no *tempo* difference' ones. While we take the observation that in 37% of crucial trials, listeners 'mixed up' the two parameters to be an important one, it does seem that an experimental design with a single tempo-related response parameter can lead to an over-estimation of the extent to which listeners interpret the manipulated dimension in terms of tempo: a proportion of responses may simply reflect listeners' identification of *difference*. Future work should explore the extent to which our Experiment 2 and 3 results, and indeed those of previous tempo perception experiments with similar designs, can be replicated using more complex designs which incorporate multiple response parameters.

One potential weakness of our experimental approach is our reliance on linear compression to achieve rate-matching. We reported in relation to Experiment 2 that clear sentence productions which required relatively little compression were most likely to be heard as slow rather than fast relative to their normal counterparts. One way to interpret this finding is that perceptual artefacts of linear compression may have informed listeners' responses to a sizeable proportion of trials. Another is that, in line with Hypothesis 2, listeners identify even small divergences from expected acoustic patterns: clear sentence productions that sounded more clearly sped up were rated as faster. Future work should therefore try to replicate our findings using approaches to controlling rate that are more faithful to original temporal structure, such as dynamic time-warping (see Beith *et al.*, in press). To disentangle the effects of clear speech cues from those of temporal compression, it would also be valuable to test whether our findings are replicated if rate-matching is achieved by slowing down normal speech, as opposed to speeding up clear speech.

Several further observations highlight listeners' potential sensitivity to fine phonetic detail which we did not control in our experimental design. In particular, one participant observed in debriefing after Experiment 2 that in many trials, 'some sound sequences' sounded particularly fast. This may relate to the low degree of coarticulation in the compressed clear productions, which makes longer consonant sequences sound particularly fast. Conversely, our own impressions of sentence production pairs in which the clear member was heard as slower was that the clear production was produced at a notably steady pace, while the corresponding normal production had more internal fluctuation in articulation rate—and therefore a higher maximum local segment rate (cf. Plug & Smith, 2021). In Experiment 3's fast tempo condition, the weak syllables in clear tokens sounded to us very saliently fast; in the slow tempo condition, some perceptually unusual segmental timing patterns gave the impression of 'slurred' articulation (cf. Koreman, 2006), although the extent of 'slurring' varied between sentences. These observations should inform further experimental work towards unpicking the relevance for tempo perception of individual acoustic correlates of speaking mode variation.

Clearly, behind the ‘H&H’ and ‘normal *vs* clear’ concepts lies a multidimensional landscape of variation, rather than a simple continuum ranging from slow hyper-articulation to rapid hypo-articulation. Speakers command many coherent styles, each locally fitted to its context, and commonly doing interactional as well as expressive and informational work. If the variation in our stimuli represents only one of many possible relations between speed and clarity, it is perhaps not surprising that listeners’ responses did not all point in the same direction for all trials.

To conclude, in natural speech, the relationship between temporal and stylistic properties reflects a host of linguistic, informational, contextual and situational factors. This study sought to tease some of these apart, exploring the impact of clear speaking style on perceived tempo, separate from the articulation rate that typically accompanies that style. We found that when articulation rate is artificially controlled by linear compression of speech, clear speech sounds, in general, faster than normal speech. This relationship is strong but not completely deterministic, reflecting other pressures in play.

ACKNOWLEDGMENTS

This research was funded by a Leverhulme Trust Research Grant (RPG-2017-060).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

Ethical approval was obtained from the University of Leeds Faculty of Arts, Humanities and Cultures Ethics Committee (LTSLCS-072).

DATA AVAILABILITY

On publication, the data associated with this study (stimuli, response analysis files) will be made openly available via the Research Data Leeds repository at https://doi.org/10.5518/1670.

SUPPLEMENTARY MATERIAL

See supplementary material at [URL will be inserted by AIP] for acoustic results and statistical analyses for the Experiment 1 stimuli.

REFERENCES

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. K. (**2020**). "Gorilla in our midst: An online behavioral experiment builder," Behav. Res. Methods **52**, 388-407.

Aylett, M., and Turk, A. (**2004**). "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang. Speech **47**, 31-56.

Bates, D., Maechler, M., Bolker, B. M., and Walker, S. C. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1-48.

Beith, A., Barr, D., and Smith, R. (**in press**). "Preserving prosody in temporal distortions of speech," in *Rhythms of speech and language*, edited by L. Meyer, and A. Strauss (Cambridge University Press, Cambridge).

Boersma, P., and Weenink, D. (**2017**). "Praat: Doing phonetics by computer (version 5.3.51) [computer program]," Retrieved from https://www.fon.hum.uva.nl/praat/.

Boltz, M. G. (**2011**). "Illusory tempo changes due to musical characteristics," Music Percept. **28**, 367-386.

Bosker, H. R., and Reinisch, E. (**2017**). "Foreign languages sound fast: Evidence from implicit rate normalization," Front. Psychol. **8**.

Bosker, H. R., Reinisch, E., and Sjerps, M. J. (**2017**). "Cognitive load makes speech sound fast, but does not modulate acoustic context effects," J. Mem. Lang. **94**, 166-176.

Bradlow, A. R. (**n.d.**). "SpeechBox," Retrieved from https://speechbox.linguistics.northwestern.edu.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255-272.

Cangemi, F. (2015). "Mausmooth (version 1.0) [computer program]," Retrieved from http://phonetik.phil-fak.uni-koeln.de/fcangemi.html.

Collins, B., and Mees, I. (**2013**). *Practical phonetics and phonology: A resource book for students*. 3rd edition. Abingdon: Routledge.

Feldstein, S., and Bond, R. N. (**1981**). "Perception of speech rate as a function of vocal intensity and frequency," Lang. Speech **24**, 387-394.

Ferguson, S. H., and Morgan, S. D. (**2018**). "Talker differences in clear and conversational speech: Perceived sentence clarity for young adults with normal hearing and older adults with hearing loss," J. Speech. Lang. Hear. R. **61**, 159-173.

Gibbon, D., Klessa, K., and Bachan, J. (**2015**). "Duration and speed of speech events: A selection of methods," Lingua Posnaniensis **56**, 59-83.

Grosjean, F. (**1977**). "Perception of rate in spoken and sign languages," Percept. Psychophys. **22**, 408-413.

Grosjean, F., and Lane, H. (**1974**). "Effects of two temporal variables on listeners' perception of reading rate," J. Exp. Psychol. **102**, 893-896.

Hazan, V., and Baker, R. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**, 2139-2152.

Hazan, V., Tuomainen, O., Kim, J., Davis, C., Sheffield, B., and Brungart, D. (**2018**). "Clear speech adaptations in spontaneous speech produced by young and older adults," J. Acoust. Soc. Am. **144**, 1331-1346.

Kisler, T., Reichel, U. D., and Schiel, F. (**2017**). "Multilingual processing of speech via web services," Comput. Speech Lang. **45**, 326–347.

Kohler, K. J. (**1986**). “Parameters of speech rate perception in German words and sentences: Duration, f0 movement, and f0 level,” Lang. Speech **29**, 115-139.

Koreman, J. (**2006**). “Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech,” J. Acoust. Soc. Am. **119**, 582-596.

Krause, J. C., and Braida, L. D. (**2002**). “Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility,” J. Acoust. Soc. Am. **112**, 2165-2172.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2017**). “LmerTest package: Tests in linear mixed effects models,” J. Stat. Softw. **82**, 1-26.

Lenth, R. (**2024**). “Emmeans: Estimated marginal means, aka least-squares means.” R package version 1.10.3099, https://rvlenth.github.io/emmeans/

Lindblom, B. (**1990**). “Explaining phonetic variation: A sketch of the H & H theory,” in *Speech production and speech modelling*, edited by W. J. Hardcastle, and A. Marchal (Kluwer Academic, Amsterdam), pp. 403-439.

Lindblom, B. (**1996**). “Role of articulation in speech perception: Clues from production,” J. Acoust. Soc. Am. **99**, 1683-1692.

Matthies, M., Perrier, P., Perkell, J. S., and Zandipour, M. (**2001**). “Variation in anticipatory coarticulation with changes in clarity and rate,” J. Speech. Lang. Hear. R. **44**, 340-353.

Ogden, R. (**2009**). *An introduction to English phonetics*. Edinburgh: Edinburgh University Press.

Pfitzinger, H. (**1999**). “Local speech rate perception in German speech,” in *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco), pp. 893-896.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). “Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech,” J. Speech. Lang. Hear. R. **28**, 96-103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). “Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,” J. Speech. Lang. Hear. R. **29**, 434-446.

Plug, L., Lennon, R., and Gold, E. (**2021**). “Articulation rates’ inter-correlations and discriminating powers in an English speech corpus,” Speech Commun. **132**, 40-54.

Plug, L., Lennon, R., and Smith, R. (**2022**). “Measured and perceived speech tempo: Comparing canonical and surface articulation rates,” J. Phonetics **95**, 101193.

Plug, L., and Smith, R. (**2021**). “The role of segment rate in speech tempo perception by English listeners,” J. Phonetics **86**, 101040.

R Development Core Team (**2008**). “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria.

Reinisch, E. (**2016**). “Natural fast speech is perceived as faster than linearly time-compressed speech,” Atten. Percept. Psychophys. **9**, 9.

Rietveld, A. C. M., and Gussenhoven, C. (**1987**). “Perceived speech rate and intonation,” J. Phonetics **15**, 273-285.

Searl, J., and Evitts, P. M. (**2013**). “Tongue-palate contact pressure, oral air pressure, and acoustics of clear speech,” J. Speech. Lang. Hear. R. **56**, 826-839.

Smiljanić, R., and Bradlow, A. R. (**1999**). “Speaking and hearing clearly: Talker and listener factors in speaking style changes,” Lang. Linguist. Compass **3**, 236-264.

Smiljanić, R., and Bradlow, A. R. (**2008**). “Temporal organization of English clear and conversational speech,” J. Acoust. Soc. Am. **124**, 3171-3182.

Smiljanic, R., and Gilbert, R. C. (**2017**). “Intelligibility of noise-adapted and clear speech in child, young adult, and older adult talkers,” J. Speech. Lang. Hear. R. **60**, 3069-3080.

Tang, K., and Shaw, J. A. (**2021**). “Prosody leaks into the memories of words,” Cognition **210**, 104601.

Vitela, A. D., Warner, N., and Lotto, A. (**2013**). "Perceptual compensation for differences in speaking style," Front. Psych. **4**.

Weirich, M., and Simpson, A. P. (**2014**). "Differences in acoustic vowel space and the perception of speech tempo," J. Phonetics **43**, 1-10.

White, L., Mattys, S. L., and Wiget, L. (**2012**). "Language categorization by adults is based on sensitivity to durational cues, not rhythm class," J. Mem. Lang. **66**, 665-679.