Perceptual learning of time-compressed and natural fast speech

Patti Adank[a)]
School of Psychological Sciences, University of Manchester, Manchester, United Kingdom and Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Nijmegen, The Netherlands

Esther Janse
Utrecht Institute of Linguistics, OTS, Utrecht University, Utrecht, The Netherlands and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 18 February 2009; revised 29 June 2009; accepted 6 August 2009)

Speakers vary their speech rate considerably during a conversation, and listeners are able to quickly adapt to these variations in speech rate. Adaptation to fast speech rates is usually measured using artificially time-compressed speech. This study examined adaptation to two types of fast speech: artificially time-compressed speech and natural fast speech. Listeners performed a speeded sentence verification task on three series of sentences: normal-speed sentences, time-compressed sentences, and natural fast sentences. Listeners were divided into two groups to evaluate the possibility of transfer of learning between the time-compressed and natural fast conditions. The first group verified the natural fast before the time-compressed sentences, while the second verified the time-compressed before the natural fast sentences. The results showed transfer of learning when the time-compressed sentences preceded the natural fast sentences, but not when natural fast sentences preceded the time-compressed sentences. The results are discussed in the framework of theories on perceptual learning. Second, listeners show adaptation to the natural fast sentences, but performance for this type of fast speech does not improve to the level of time-compressed sentences.
 [DOI: 10.1121/1.3216914]

PACS number(s): 43.71.Es, 43.71.Bp, 43.71.Gv [PEI] Pages: 1–XXXX

I. INTRODUCTION

Within a given conversation, speakers often vary their speech rate considerably (1984), ranging between 140 and 180 words/min. These on-line changes in speaking rate affect qualitative aspects of speech: at higher rates, speech is produced with generally more coarticulation and assimilation (Browman and Goldstein, 1990; Byrd and Tan, 1996) sometimes even leading to deletion of segments (Ernestus *et al.*, 2002; Koreman, 2006). Moreover, people increase their speech rate in a nonlinear fashion: higher speaking rates generally affect consonant durations less than vowel durations (Lehiste, 1970; Max and Caruso, 1997). In addition, durations of unstressed syllables in polysyllabic words are reduced more than stressed syllables (Peterson and Lehiste, 1960). These phonetic and phonological consequences of the variations in speaking rate pose a potential problem for listeners, forcing them to constantly normalize for varying speech rate (Green *et al.*, 1994; Miller *et al.*, 1984a; Miller and Liberman, 1979).

Apart from these latter studies on local rate effects on phonetic perception of specific phoneme contrasts, there is a body of research on more gradual adaptation to artificially time-compressed speech. Artificial time compression is a method for artificially shortening the duration of an audio signal without affecting the fundamental frequency of the signal (Golomb *et al.*, 2007; Pallier *et al.*, 1998; Sebastián-Gallés *et al.*, 2000; Wingfield *et al.*, 2003). Listeners can adapt to sentences compressed up to 38% of their original duration within 10–20 sentences (Dupoux and Green, 1997). Adaptation to this manipulation is not immediate, but takes place during exposure to a number of sentences that are initially of very poor intelligibility. While adaptation to time-compressed speech has provided useful insights on general adaptation processes in speech comprehension, it is questionable whether time-compressed speech itself provides a useful model for adaptation to the specific characteristics of naturally produced fast speech. First of all, there is evidence that natural fast speech is more difficult to process than speech that is artificially time compressed to the same rate (Janse, 2004). Second, modern time-compression algorithms (Moulines and Charpentier, 1990) do not significantly affect the long-term spectral characteristics of the original speech signal, while allowing for careful manipulation of the temporal characteristics. Natural fast speech, on the other hand, differs from speech delivered at a normal speaking rate in both spectral and temporal characteristics (Koreman, 2006; Wouters and Macon, 2002).

The aims of the present study were twofold. First, we wanted to establish if listeners adapt to naturally produced fast speech and if so, how this adaptation process compares to adaptation to time-compressed speech. While adaptation to time-compressed speech is usually determined with participants reporting keywords in a sentence (Dupoux and Green, 1997; Golomb *et al.*, 2007; Pallier *et al.*, 1998), adaptation in the present study was measured using reaction times and percent correct as (i) they were expected to pro-

[a)]Author to whom correspondence should be addressed. Electronic mail: patti.adank@manchester.ac.uk

vide a more fine-grained measure than percent correctly reported keywords only and (ii) to avoid ceiling effects in the time-compressed speech condition. Speech can be highly time compressed before identification scores of listeners drop below ceiling level. Such fast rates of speech can hardly be attained by humans speeding up their speech rate, which makes it difficult to compare the two types of fast speech at the same rate. Also, Clarke and Garrett (2004) used reaction times to show adaptation to foreign-accented speech. We therefore used a speeded sentence verification task to monitor the adaptation process. This task is based on the Speech and Capacity of Language Processing Test, or SCOLP (Baddeley *et al.*, 1992) of which an aural version was previously described in May *et al.*, 2001; Adank *et al.*, 2009. SCOLP is originally a written test in which the participant verifies as many sentences as possible in 2 min. The sentences are all obviously true or false and all consist of a mismatch of subject and predicate from true sentences (e.g., *Tomato soup is a liquid* versus *Tomato soup is people*). Overall, it provides a sensitive and reliable measure of the speed of language comprehension. When transformed to a speeded verification task, it can be used to determine the cognitive processing cost of a specific task or process, as demonstrated by Adank *et al.* (2009), who used the task to determine the relative cognitive load of comprehension of regionally accented sentences versus sentences in the standard language in noise. A decrease in the speed of processing after exposure to time-compressed speech can thus be taken to signal perceptual learning of the acoustic consequences of, for instance, time compressed or naturally fast speech. We created a Dutch version of the SCOLP sentences as the experiment was run in The Netherlands, with Dutch listeners. Like the English version, the Dutch version was made up of sentences that consisted of a noun plus predicate, like the SCOLP sentences. A total of 90 sentence pairs were constructed (90 true and 90 false). For example, "Tomaten groeien aan planten" (*tomatoes grow on plants*) as a true sentence and "Tomaten hebben sterke tanden" (*tomatoes have strong teeth*) as a false sentence. All sentences of the Dutch version designed for the present study are listed in Table I.

Second, we aimed to establish whether there is transfer of learning in the adaptation process between naturally fast and time-compressed speech: does exposure to time-compressed speech before being exposed to naturally fast speech affect the adaptation process (and vice versa)? Transfer of learning involves the application of skills or knowledge learned in one context to another context (Cormier and Hagman, 1987; Haskell, 2001; Thorndike and Woodforth, 1901). Transfer of learning has been found in the auditory domain for nonspeech stimuli (Delhommeau *et al.*, 2005; Delhommeau *et al.*, 2002) and speech stimuli (Bradlow and Bent, 2008; McClaskey *et al.*, 1983; Tremblay *et al.*, 1997). Transfer of learning was, for instance, reported for auditory frequency discrimination tasks: Delhommeau *et al.* (2002) measured listeners' frequency discrimination thresholds (FDTs) (the smallest audible difference frequency, Δf, around a center frequency) for four center frequencies (750, 1500 3000, and 6000 Hz) before and after training. Listeners were then trained for a specific center frequency (e.g., 750 Hz) and then subsequently tested again at all four center frequencies. Delhommeau *et al.* (2002) found that training at a specific frequency lowered FDTs for that frequency and that the improvement transferred to the other (untrained) frequencies. Furthermore, McClasky *et al.* (1983) trained listeners to perceive prevoiced labial syllables and found that they generalized their newly learned ability to prevoiced alveolar syllables, while Tremblay *et al.* (1997) found a preattentive effect signaling transferred learning on listeners discriminating between prevoiced alveolar stops after having been trained to discriminate prevoiced labial stops. Finally, transfer of learning has been found for adaptation to a foreign accent across speakers (Bradlow and Bent, 2008). Bradlow and Bent (2008) found that listeners were better able to comprehend sentences in a foreign accent spoken by a novel speaker after having adapted to other speakers with the same foreign accent. In the present experiment, we tested whether having adapted to one type of fast speech facilitates adaptation and/or general performance for the other type. Time-compressed sentences differ from normal sentences only in their temporal characteristics, while natural fast sentences differ from normal sentences in their temporal characteristics as well as their spectral characteristics. Transfer of learning between the two speech types, and between temporal and spectral variations will be tested in the present experiment using a between-subjects design with two listener groups in which the order of presentation of the two speech types is varied. Both groups first verified 60 sentences spoken at a normal rate. During this normal-rate block, listeners could get used to the task and the type of sentences. Subsequently listeners in group 1 listeners verified 60 natural fast sentences before finally verifying 60 time-compressed sentences, while listeners in group 2 first verified 60 time-compressed sentences followed by 60 natural fast sentences. This division into two groups allowed us to study the effect of the type of compression (artificial or natural) on the adaptation process and to test whether there is transfer of learning. If there is transfer of learning from time-compressed speech to natural fast speech, then performance (i.e., accuracy) for the natural fast speech should be higher for group 2 than for group 1. Alternatively, if there is transfer of learning from natural fast speech to time-compressed speech, then performance for the time-compressed sentences should be higher for group 1 than for group 2. Figure 1 presents an overview of the order in which the three speech types were presented to both groups.

II. METHOD

A. Participants

Forty-two participants (nine male, mean age 22.1, stddev. 4.3 years, median age 22 years, range 18–41 years) took part in the study. All were native speakers of Dutch from The Netherlands, with no history of oral or written language impairment, or neurological or psychiatric disease. None reported any hearing problems or any previous experience with time-compressed speech. Listeners were randomly

AQ: #6

TABLE I.

No.	True	False
1	Makrelen ademen door kieuwen	Chirurgen groeien aan planten
2	Bevers bouwen dammen in de rivier	Bevers groeien in een moestuin
3	Bisschoppen dragen kleren	Wortels hebben een beroep
4	Ezels dragen zware vrachten	Bromfietsen hebben een snavel
5	Pinguns eten veel vis	Slagers hebben een staart
6	Tomaten groeien aan planten	Forellen hebben een vacht
7	Wortels groeien in een moestuin	Haaien hebben handen
8	Architecten hebben een beroep	Nachtegalen hebben manen
9	Roodborstjes hebben een snavel	Pinguns hebben schubben
10	Tijgers hebben een start	Tomaten hebben sterke taanden
11	Luipaarden hebben een vacht	Makrelen hebben veren
12	Vaders hebben handen	Lepels hebben vier poten
13	Leeuwen hebben manen	Schuurtjes hebben voelsprieten
14	Forellen hebben schubben	Aardappels hebben voeten
15	Haaien hebben sterke tanden	Leeuwen hebben winkels
16	Nachtegalen hebben veren	Mieren zijn van hout
17	Beren hebben vier poten	Vlinders komen van schapen
18	Vlinders hebben voelsprieten	Hamers kruipen op hun buik
19	Wetenschappers hebben voeten	Auto's kunnen goed zwemmen
20	Slagers hebben winkels	Tantes kunnen in winkels gekocht worden
21	Kasten zijn van hout	Kroketten kunnen koppig zijn
22	Lammetjes komen van schapen	Asperges kunnen ver vliegen
23	Ratelslangen kruipen op hun buik	Messen zijn eetbaar
24	Otters kunnen goed zwemmen	Biefstukken moeten lang studeren
25	Blikopeners kunnen in winkels gekocht worden	Wijnflessen rijden op de weg
26	Ezels kunnen koppig zijn	Wandelschoenen vliegen rond op zoek nar voedsel
27	Ganzen kunnen ver vliegen	Luipaarden voeren het bevel op scheppen
28	Druiven zijn eetbaar	Roodborstjes werken in de politiek
29	Chirurgen moeten lang studeren	Ezels wonen in een klooster
30	Bromfietsen rijden op de weg	Ratelslangen worden gebruikt als keukengerei
31	Bijen vliegen rond op zoek naar voedsel	Presidenten worden gebruikt voor het eten van soep
32	Kapiteins voeren het bevel op schepen	Kapiteins worden gebruikt voor opslag
33	Presidenten werken in de politiek	Monniken worden geschild
34	Monniken wonen in een klooster	Tijgers worden gemaakt in een fabriek
35	Messen worden gebruikt als keukengerei	Taarten worden in de tuin gebruikt
36	Lepels worden gebruikt voor het eten van soep	Architecten worden verkocht door slagers
37	Schuurtjes worden gebruikt voor opslag	Politieagenten hebben een kurk
38	Aardappels worden geschild	Heggenscharen zijn altijd vrouwen
39	Sloffen worden gemaakt in een fabriek	Ezels zijn deel van de familie
40	Heggenscharen worden in de tuin gebruikt	Giraffes zijn fruit
41	Biefstukken worden verkocht door slagers	Wetenschappers zijn gefabriceerde goederen
42	Wijnflessen hebben een kurk	Beren zijn gefrituurd
43	Tantes zijn altijd vrouwen	Ganzen zijn groenten
44	Ooms zijn deel van de familie	Ministers worden in een oven gebakken
45	Bananen zijn fruit	Olifanten zijn klein
46	Wandelschoenen zijn gefabriceerde goederen	Kasten zijn levende wezens
47	Kroketten zijn gefrituurd	Kakkerlakken zijn meubels
48	Asperges zijn groenten	Ooms zijn om op te zitten
49	Taarten worden in een oven gebakken	Dolfijnen gebruiken benzine
50	Mieren zijn klein	Sloffen zijn insecten
51	Olifanten zijn levende wezens	Bananen zijn zoogdieren
52	Tafels zijn meubels	Vaders zitten in de gereedschapskist
53	Stoelen zijn om op te zitten	Lammetjes zitten in de regering
54	Auto's gebruiken benzine	Bijen hebben een lange nek
55	Kakkerlakken zijn insecten	Stoelen lopen op straat
56	Dolfijnen zijn zoogdieren	Een kameel is een soort vogel
57	Hamers zitten in de gereedschapskist	Een panter heeft vleugels
58	Ministers zitten in de regering	Een kool is een soort vrucht
59	Giraffes hebben een lange nek	Een boon is zoet
60	Politieagenten lopen op straat	Een mus is een zoogdier

TABLE I. (*Continued.*)

No.	True	False
61	Een pelikaan is een soort vogel	Een overhemd is een lichaamsdeel
62	Een adelaar heeft vleugels	Een schoen heeft vingers
63	Een aardbei is een soort vrucht	Een aap is een soort vis
64	Een appel is zoet	Een boor is een muziekinstrument
65	Een varken is een zoogdier	Een viool is een werktuig
66	Een been is een lichaamsdeel	Mensen dragen een broek aan hun handen
67	Een hand heeft vingers	Sommige mensen hebben giraffes als huisdier
68	Een stekelbaars is een soort vis	De meeste auto's rijden op appelsap
69	Een gitaar is een muziekinstrument	Denemarken is een land in Afrika
70	Een waterpomptang is gereedschap	Een paard heeft drie benen
71	Mensen dragen sokken aan hun voeten	Roken is goed voor je gezondheid
72	Sommige mensen hebben honden als huisdier	Een uur is vijfenveertig minuten
73	De meeste vrachtwagens rijden op diesel	Melk bevat alcohol
74	Spanje is een land in Europa	Mensen hebben op de zon gelopen
75	Een paard heeft vier benen	Sommige mensen drinken thee met zout
76	Beweging is goed voor je gezondheid	Olifanten eten soms mensen op
77	Een minuut heeft zestig seconden	Een boom heeft melk nodig om te leven
78	Bier bevat alcohol	Een groen licht betekent stop
79	Mensen hebben op de maan gelopen	Papier wordt gemaakt van onkruid
80	Sommige mensen drinken koffie met suiker	Een fiets is een oorlogwapen
81	Krokodillen eten soms kinderen op	Boeddhisme is een politieke theorie
82	Een plant heeft water nodig om te leven	Spaghetti is een Frans gerecht
83	Een rood licht betekent stop	Een loodgieter kan je helpen als je ziek bent
84	Perkament wordt gemaakt van leer	Fietsen is meestal langzamer dan lopen
85	Een tank is een oorlogwapen	Kinderen zijn nooit bang in het donker
86	Baksteen is een goed materiaal voor gebouwen	Een schip is een soort meubel
87	Boekhouden is een beroep	Een sinaasappel is knapperig
88	Juni is een zomermaand	Een baksteen is een edelsteen
89	Een step is goed te besturen	De hoofdstad van Nederland is Brussel
90	Een vrachtwagen heeft een motor	Een kip kan goed gitaar spele

allocated to the two groups: 21 to group 1 and 21 to group 2. All gave written informed consent and were paid for their participation or received course credit.

B. Speech material

Recordings were made of a 31-year-old male speaker of Standard Dutch who had lived in The Netherlands all his life. Recordings were made of two versions of the 180 sentences listed in Table I. The procedure for the sentences produced at a normal rate was as follows. First, the sentence was presented on the computer screen in front of the speaker. He was instructed to first quietly read the sentence and to subsequently pronounce the sentence as a declarative statement at his normal speech rate. All sentences were recorded once. Next, the natural fast sentences were recorded. A sentence was presented on the computer screen. Again the speaker was asked to first read the sentence in silence. After that he produced the sentence four times in quick succession, as it was found that this was the best way for him to produce the sentences as fast and fluently as possible. The recordings were made in a sound-treated room, using a Sennheiser ME64 microphone, which was attached to an Alexis Multimix USB audio mixing station. The recordings were saved at 44 100 Hz to hard disk directly via an Imix DSP chip plugged into the Alexis Multimix and to the USB port of an Apple Macbook. PRAAT (Boersma and Weenink, 2003) was used to save all sentences into separate sound files with begin and end trimmed at zero crossings (trimming on or as closely as possible to the onset and offset of initial and final speech sounds) and resampled from 44 100 to 22 050 Hz. For the natural fast sentences, in the great majority of cases (>95%), the second sentence was selected out of the quartet of sentences recorded, as these were judged by the experimenters to be the best examples (fastest as well as most fluent). Subsequently, the durations of the 2×180 sentences used in the experiment were calculated. The normal speech rate sentences consisted of 4.7 (intended) syllables on average (range 3–12 syllables, stdev. 0.6 syllables) and the speech rate of the natural fast sentences was 10.2 syllables/s (stddev. 1.6 syllables). On average, the selected natural fast sentences were pronounced at 46.0% of the duration of the normal speech rate sentences, with the fastest item pro-

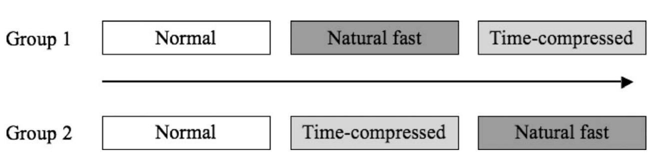

FIG. 1. Overview of the experimental design. Group 1 (top) was presented first with 60 normal sentences, immediately followed by 60 natural fast sentences, and followed by 60 time-compressed sentences. Group 2 (bottom) was presented first with 60 normal sentences, immediately followed by 60 time-compressed sentences, and followed by 60 natural fast sentences

nounced at 32.6% and the slowest at 88.7%. Next, the time-compressed sentences were obtained by digitally shortening them with PSOLA (Pitch Synchronous Overlap and Add) (Moulines and Charpentier, 1990), as implemented in PRAAT. Compression rates were established per sentence: each individual time-compressed sentence was matched in rate to its corresponding natural fast item. For instance, if a natural fast sentence was pronounced at 48% of the duration of the normal speed sentence (i.e., twice as fast), then the compression rate for the PSOLA version of that sentence was set to 48%. Subsequently, the normal sentences and the natural fast sentences were all resynthesized at 100% of their original duration using PSOLA. Finally, the intensity of each of the 540 (180 sentences $\times$ 3 variants) sound files was peak normalized at 99% of its maximum amplitude and scaled to 70 dB sound pressure level.

C. Procedure

All listeners were tested individually in a sound-treated booth and received written instructions. Responses were made using a button box with the index finger (true responses) and middle (false response) finger of their dominant hand. The stimuli were presented over Sennheiser HD477 headphones at a comfortable sound level per participant. Stimulus presentation and response time (RT) measurement were performed using PRESENTATION (Neurobehavioral Systems, Albany, CA). Response times were measured relative to the end of the audio file, following May *et al.* (2001) and Adank *et al.* (2009).

Each trial proceeded as follows. First, the stimulus sentence was presented. Second, the program waited for 3 s before playing the next stimulus, allowing the participant to respond. If the participant did not respond within 3 s, the trial was recorded as *no response*. Participants were asked to respond as quickly and accurately as possible and they were told that they did not have to wait until the sentence was finished (allowing for negative RTs, as RT was calculated from the offset of the sound file). Six familiarization trials were presented prior to the start of the experiment. The familiarization sentences had been produced by the same speaker and were spoken at a normal speech rate. The familiarization sentences were not included in the actual experiment. The test sentences were presented in a semirandomized order per participant and true and false sentences were counterbalanced across experimental blocks. Within an experimental condition, no true-false sentence pairs were presented. For instance, the true and false versions of sentence 2 ("Bevers bouwen dammen in de rivier" (English: *Beavers build dams in the river*) and "Bevers groeien in een moestuin" (English: *Beavers grow in the vegetable patch*), see Table I), were never presented within one experimental condition. Total duration of the listening study was 15 min, without breaks.

III. RESULTS

The data from one of the participants of group 1 were excluded from the analysis, as her average RTs were more than two standard deviations slower than the average across all participants. Due to a programming error, six participants (three per listener group) got 70 (instead of 60) time-compressed sentences and they then got 50 (instead of 60) natural fast sentences. We excluded the last ten time-compressed trials for these participants and recoded trial number within the natural fast block of sentences.

Figure 2 and Table II show the average error percentages for both groups per speech type for the data grouped into ten subsequent miniblocks of sentences, in order to see adaptation over exposure time. Likewise, Fig. 3 and Table III show average RTs for the two groups (in milliseconds, measured from sentence offset) for the three speech types, again broken down in ten (mini)blocks of six sentences. The results in Figs. 2 and 3 are only plotted in ten miniblocks of six subsequent sentences for demonstration purposes. The statistical analysis was performed with trial as a continuous variable.

The results were analyzed with linear mixed effects models with participant and item as crossed random effects (Pinheiro and Bates, 2000; Quené and van den Bergh, 2004). One model was fitted to the binomial accuracy data (a response being correct or incorrect), and one model was fitted to the RT data (for correct responses only). Order was a between-participant factor, and speech type (normal, time compressed, or natural fast) and trial (within each block of 60 sentences of that particular speech type) were within-participant factors. As mentioned above, we chose to look for effects of trial to study adaptation, rather than of miniblock (see Figs. 2 and 3) because trial provided us with the most fine-grained continuous variable in relation to adaptation (note, though, that an alternative analysis with the variable miniblock, instead of trial, produced highly similar results). We also entered the (within-participants and between-items) factor of whether the sentence ought to elicit a true or a false response because participants may have found it easier to verify either type. Systematic stepwise model comparisons using likelihood ratio tests established the best-fitting model.

A. Accuracy

The linear mixed-effects model for accuracy had as dependent variable whether or not the response was correct ($N=7320$). The within-subjects factor speech type had three levels (normal, time compressed, and natural fast). The linear mixed effects model gives as output whether each of the levels differs significantly from the one mapped onto the intercept (in this case, the normal-rate sentences). Beta values are provided for significant effects and interactions (with standard error in brackets) as well as significance levels.

Performance on the natural-fast sentences was significantly poorer than performance on the normal-rate sentences [$\beta=-2.234$ (0.382), $p<0.001$], but performance on the time-compressed sentences was not. There was an overall effect of trial [$\beta=0.040$ (0.014), $p<0.01$], indicating that performance improved over trials within speech type. The effect of correct response (true or false) also significantly affected accuracy: participants showed better performance for the false than the true sentences [$\beta=0.708$ (0.236), $p<0.01$]. Overall, the two listener groups did not differ in performance [order: $\beta=1.048$ (0.552), n.s.].

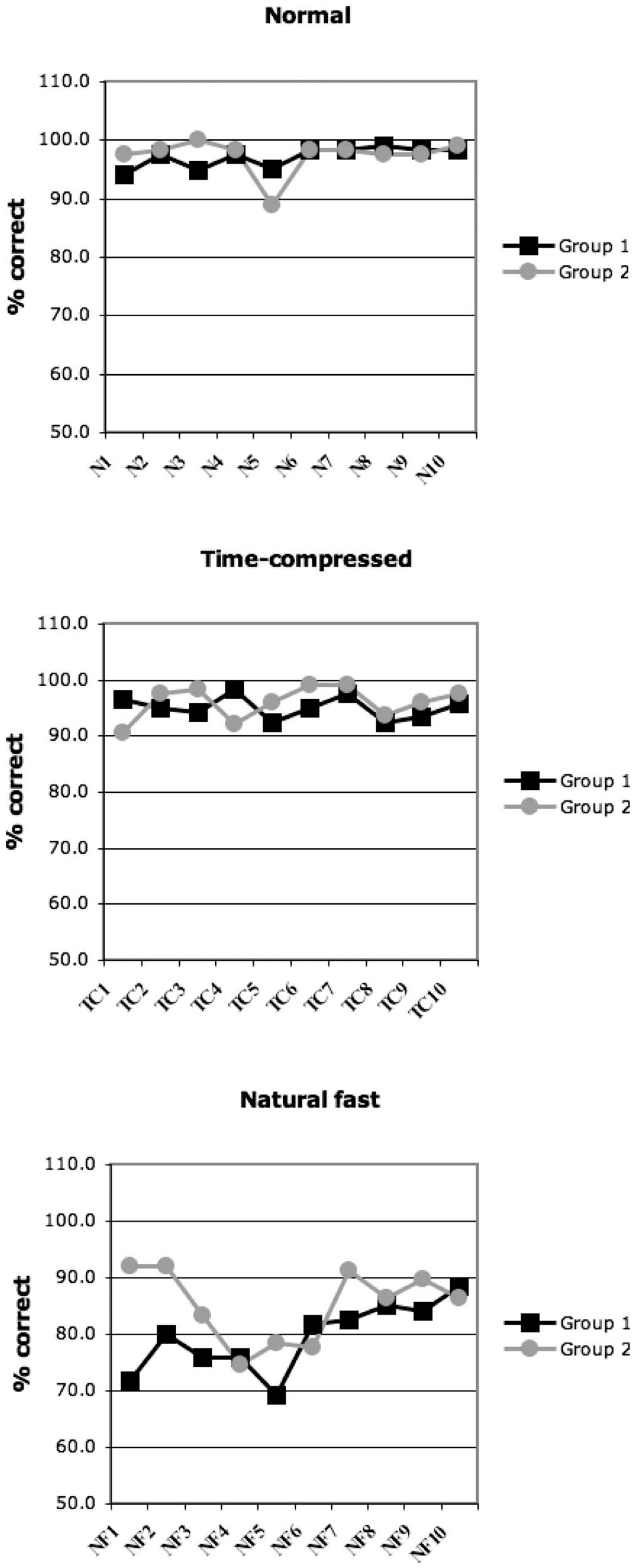

FIG. 2. Average percent correct (%) per block of six sentences correct for the normal speed condition (top panel, miniblocks N1–N10), Time-compressed condition (middle panel, miniblocks TC1–TC10), and the natural fast condition 1 (bottom panel, miniblocks NF1–NF10) for both groups (group 1 in black and group 2 in gray).

Speech type interacted with trial: for the time-compressed speech, improvement over trials was less than in the other two speech types [$\beta = -0.041$ (0.015), $p < 0.05$]. This was modified further by a three-way interaction of order by speech type by trial [$\beta = 0.0584$ (0.022), $p < 0.01$]: that there was less improvement over trials for the time-compressed speech, relative to the other speech types, was mainly the case for the listeners in group 1, who heard the time-compressed sentences after they had been presented with the natural-fast speech. It was less true for the group 2 listeners who heard the time-compressed sentences before the natural-fast sentences. This fits in with slightly poorer overall performance for group 2 on the time-compressed sentences, as suggested by an order by speech type interaction [$\beta = -1.469$ (0.707), $p < 0.05$] for the time-compressed speech.

The data were also analyzed for the three speech types separately to investigate whether there is improvement or adaptation over trials and to see whether the order in which listeners heard the conditions mattered. Bonferroni correction was applied to the outcomes of the subset analyses (we analyzed three subsets and the critical p-value of 0.05 was thus set to 0.05/3, resulting in a critical value of 0.017).

For the normal-rate sentences, there was an overall effect of trial, meaning that accuracy performance improved over trials [$\beta = 0.050$ (0.025), $p = 0.044$], but note that this does not exceed the Bonferroni-corrected critical value for significance. There was no difference between the two orders (i.e., between the two listener groups, and note that normal-rate sentences were presented first in both orders) and no interaction between trial and order. The effect of correct response (true or false) was not significant in this subset.

For the time-compressed sentences, there was no overall effect of trial and no interaction between order and trial. The only effect approaching significance was that of correct response: stating that the sentence is false being the easier response [$\beta = 1.091$ (0.483), $p = 0.024$, which does not meet the Bonferroni-corrected threshold value]. This subanalysis complements the picture provided by the two-way and three-way interactions reported above in the overall analysis. Unlike the other speech conditions, there is no improvement in accuracy over time-compressed trials (this was particularly the case if the time-compressed sentences were presented as the last speech condition, but when the time-compressed condition preceded the natural fast condition improvement over trials was not significant either).

For the natural fast sentences, there was an overall order effect [$\beta = 0.957$, (0.29), $p < 0.001$]. This shows that listeners who got this condition last (i.e., after they had been presented with the time-compressed condition) had overall higher accuracy than listeners who got this condition before the time-compressed condition. Second, there was an overall effect of trial [$\beta = 0.022$ (0.008), $p < 0.01$], indicating that accuracy improved over trials. Furthermore, there was an order by trial interaction [$\beta = -0.016$ (0.008), $p = 0.047$], showing that listeners who got the natural fast sentences last showed a smaller improvement over trials than listeners who got the natural fast sentences before the time-compressed sentences (note though that this interaction fails to reach significance if we take the Bonferroni correction into account). Finally, there was a significant effect of correct response, which means that false sentences were easier to verify than

TABLE II. Mean percent error plus standard deviations (stddev) for both groups for the block 1 three speech types for the ten blocks of six sentences.

% correct		Normal		Time compressed		Natural fast	
		Mean	Stddev	Mean	Stddev	Mean	Stddev
Group 1	Block 1	94.2	23.5	96.7	18.0	71.7	45.3
	Block 2	97.5	15.7	95.0	21.9	80.0	40.2
	Block 3	95.0	21.9	94.2	23.5	75.8	43.0
	Block 4	97.5	15.7	98.3	12.9	75.8	43.0
	Block 5	95.0	21.9	92.5	26.4	69.2	46.4
	Block 6	98.3	12.9	95.0	21.9	81.7	38.9
	Block 7	98.3	12.9	97.5	15.7	82.5	38.2
	Block 8	99.2	9.1	92.5	26.4	85.0	35.9
	Block 9	98.3	12.9	93.3	25.0	84.2	36.7
	Block 10	98.3	12.9	95.8	20.1	88.3	32.2
Group 2	Block 1	97.6	15.3	90.5	29.5	92.1	27.1
	Block 2	98.4	12.5	97.6	15.3	92.1	27.1
	Block 3	100.0	0.0	98.4	12.5	83.3	37.4
	Block 4	98.4	12.5	92.1	27.1	74.6	43.7
	Block 5	88.9	31.6	96.0	19.6	78.6	41.2
	Block 6	98.4	12.5	99.2	8.9	77.8	41.7
	Block 7	98.4	12.5	99.2	8.9	91.3	28.3
	Block 8	97.6	15.3	93.7	24.5	86.5	34.3
	Block 9	97.6	15.3	96.0	19.6	89.7	30.5
	Block 10	99.2	8.9	97.6	15.3	86.5	34.3

true sentences [β=0.561 (0.275), $p<0.05$]. This subset analysis clearly shows that both listener groups showed improvement over the course of the 60 natural fast sentences and that order mattered: the group who had already been presented with the time-compressed materials had overall better performance than the other group.

B. Response times

Figure 3 and Table III show the average RTs per speech type. The results are again plotted in ten (mini)blocks of six subsequent sentences each. The statistical analysis, as in the accuracy analysis, was performed with trial as a continuous variable. A linear mixed effect model was fitted to the RTs (measured from sentence offset) of the correct decisions (N =6716). As in the previous analysis, the linear mixed effect model gives as output whether each of the levels differs significantly from the one mapped onto the intercept (i.e., the normal-rate sentences).

Response times were significantly longer in the time-compressed condition than in the normal-rate condition [β =256 (21.5), $p<0.001$]. The same was true for the natural fast sentences [β=452 (22.7), $p<0.001$]: RTs were longer compared to the normal-rate sentences. There were no overall effects of order, trial, or correct response. Correct response did interact with speech type, however: in the natural fast sentence condition, listeners took longer to decide that sentences were false [β=239 (19.9), $p<0.001$]. Even though there was no overall trial effect, there were significant interactions between speech type and trial. In the time-compressed condition, the effect of trial differed from that in the normal-rate condition [β=−1.164 (0.547), $p<0.05$], suggesting that responses did get faster over the time-compressed trials. In the natural fast condition, the trial effect was also different from that in the normal-rate condition [β=−1.450 (0.576), $p<0.05$], suggesting that responses did get faster over the natural fast trials. None of the other interactions proved significant.

As in the accuracy analysis, RTs were also analyzed per speech condition to complement the picture of the overall analysis. Bonferroni correction was applied to the critical value for these subset analyses (0.05/3=0.017). For the normal-rate sentences, there were no significant effects of order, trial, or of correct response. There were no significant interactions either. For the time-compressed speech, there was a significant effect of trial [β=−1.534 (0.641), p =0.017, which just satisfies the Bonferroni corrected critical value]. Figure 3 shows that this speeding up of responses over trials was found mainly in the initial two-three miniblocks. There was no effect of order or of correct response. The interaction between trial and order was not significant either, indicating that listeners in both order groups got faster over trials. For the natural fast sentences, the data showed an effect of trial [β=−2.091 (1.111), p=0.060, which does not meet the criterion for significance] and of correct response [β=214.9 (40.79), $p<0.001$]. There was no interaction between order and trial, which means that both groups tended to become somewhat faster over trials.

The results for the time-compressed speech replicate results from Clarke and Garrett (2004), who found that listener got faster at a RT task after presentation of a small number of sentences. Our results show that group 1 got 185 ms faster between the first and the second miniblock of six sentences, while group 2 became 84 ms faster.

In sum, the RT analysis clearly confirms the difficulty

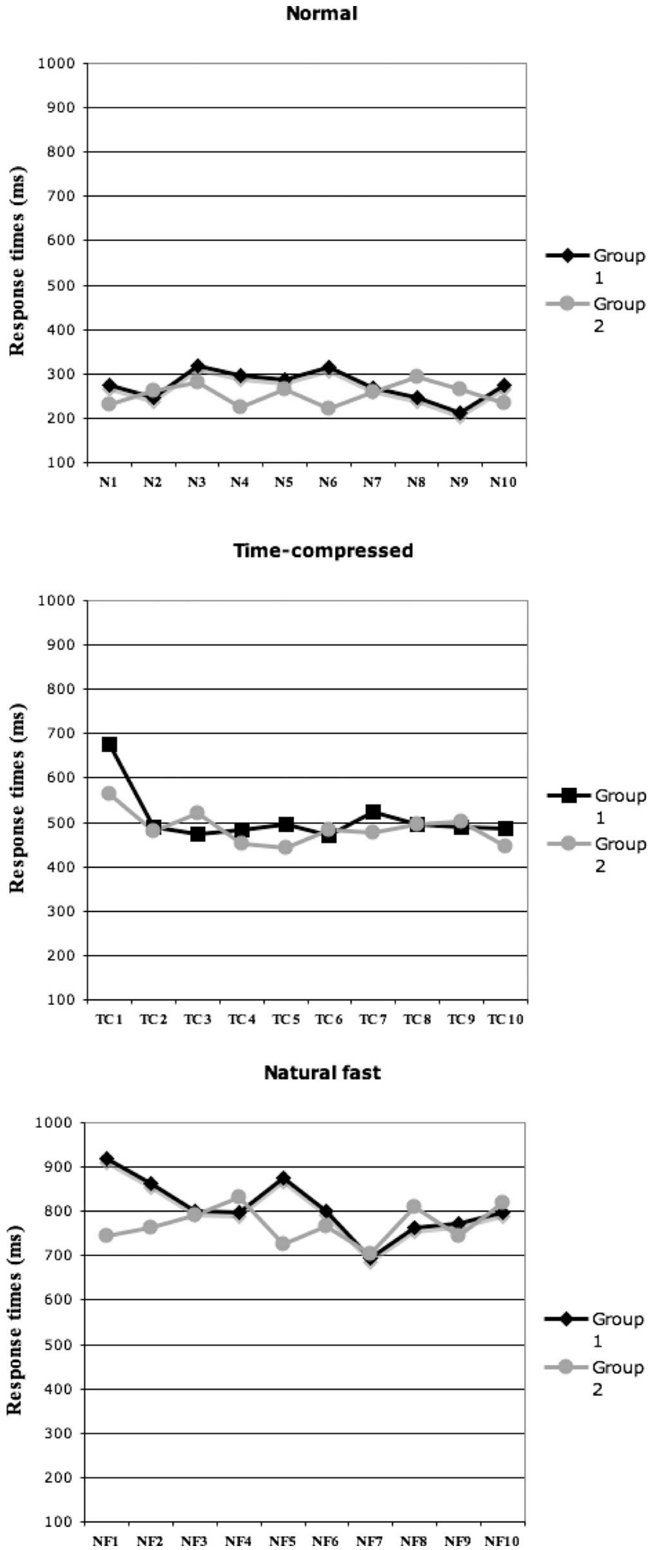

FIG. 3. Average RTs in millisecond per block of six sentences correct for the normal speed condition (top panel, miniblocks N1–N10), time-compressed condition (middle panel, miniblocks TC1–TC10), and the natural fast condition 1 (bottom panel, miniblocks NF1–NF10) for both groups (group 1 in black and group 2 in gray).

hierarchy of the three speech types also seen in the accuracy scores: listeners were fastest to respond to the normal-speed sentences, slower for the time-compressed sentences, and slowest for the natural fast sentences. The RTs were not affected by the order in which the two fast speech types were presented. Importantly, whereas adaptation to time-compressed speech did not show up as improved accuracy over trials, it was found in decreased RTs over trials. Adjustment to natural fast speech was found both in improved accuracy and in somewhat decreased RTs over trials.

Finally, one should note that any learning observed in the normal-rate condition indicates that participants needed more sentences than the six sentences in the familiarization block to get used to the task of sentence verification. Even if accuracy over the first half (30) of the normal-rate sentences is compared to accuracy in the second half, performance is significantly better in the second half. The importance of ruling out rival explanations (such as practice effects) for improved performance over trials has always been an issue in adaptation studies (Clarke and Garrett, 2004; Dupoux and Green, 1997).

IV. GENERAL DISCUSSION

We sought to establish whether listeners learn to adapt to naturally fast speech and if so, how this process compares to learning to adapt to time-compressed speech. Two groups of listeners participated in a speeded sentence verification experiment. Both groups first verified a series of sentences at a normal speaking rate. Subsequently, listeners in group 1 verified a series of natural fast sentences, followed by a series of time-compressed sentences, while this order was reversed for group 2.

The results have shown three important points. First, listeners adapt to natural fast speech. Gradual adaptation had been shown for artificially time-compressed speech materials, but not yet for natural fast speech. Natural fast speech involves a greater spectrotemporal deviation from a normal-rate speech signal than artificial time compression. Listeners' performance clearly showed that natural-fast speech is more difficult to process than artificially time-compressed speech due to the greater spectrotemporal variation, as was previously shown in Janse, 2004. The present finding that listeners are nevertheless able to adapt to natural fast speech complements the earlier findings of adaptation to highly compressed speech.

The second important point is that we have shown transfer of learning from adaptation to time-compressed speech to naturally produced fast speech. The group who had been presented with time-compressed material *before* they were presented with the natural fast material (group 2) showed generally higher accuracy for the natural fast materials. Listeners in this group benefited from having already adapted to the temporal manipulation—the time-compressed sentences—before being presented with sentences that showed temporal compression as well as spectral variation. Furthermore, their adaptation curve was shallower, because they started off higher, than that of the group who got the natural fast sentences first.

Third, whether there is transfer of learning from the natural fast speech to the time-compressed condition was less clear. One could argue that if listeners had adapted to natural fast speech, which involves a fast rate and greater

TABLE III. Mean RTs in ms plus standard deviations (Stddev) for both groups for the three speech types for the ten blocks of six sentences.

RT (ms)		Normal		Time compressed		Natural fast	
		Mean	Stddev	Mean	Stddev	Mean	Stddev
Group 1	Block 1	231	304	676	409	918	528
	Block 2	261	395	491	301	863	508
	Block 3	280	290	474	303	800	484
	Block 4	223	266	482	312	799	442
	Block 5	264	296	496	340	876	502
	Block 6	223	341	471	383	800	480
	Block 7	257	316	524	413	695	371
	Block 8	292	387	496	334	762	453
	Block 9	266	333	490	302	771	489
	Block 10	232	299	485	316	796	497
Group 2	Block 1	231	304	565	381	745	493
	Block 2	261	395	481	346	762	516
	Block 3	280	290	520	365	791	477
	Block 4	223	266	453	278	832	546
	Block 5	264	296	443	288	727	440
	Block 6	223	341	482	341	767	471
	Block 7	257	316	478	273	703	446
	Block 8	292	387	495	303	809	556
	Block 9	266	333	501	308	744	460
	Block 10	232	299	446	335	818	513

spectral smearing, time-compressed speech ought to be relatively easy. Our results do not confirm this argument, however. Both groups showed adaptation to artificial time compression in terms of decreased RTs over trials and there was no evidence for a difference in slope. Apparently, transfer of learning shows up more clearly if one is presented with speech conditions of increasing complexity rather than if the most difficult condition is followed by an easier condition.

The present study replicated the effect of learning on reaction times (Clarke and Garrett, 2004) for the time-compressed speech. Participants became faster but not more accurate for the time-compressed sentences. However, they became more accurate and somewhat faster for the natural fast sentences. This difference between time-compressed and natural fast speech may be explained by the overall difficulty of the two speech types: listeners in both groups made more errors and showed longer RTs for the natural fast sentences than for the time-compressed sentences. After presentation of approximately 30 sentences, they were able to understand the sentences better, but they still needed longer processing to perform the task adequately.

In the experiment, the time-compression factor varied per stimulus. The sentences in the time-compressed condition were matched in compression factor with the natural fast sentences. It is unclear how this may have affected the extent to which participants adapted to the manipulation. There is some evidence that phonetic variability during exposure/training aids perceptual learning (Logan *et al.*, 1991). However, one study on adapting to time-compressed speech shows that a change in compression rate can lead to a temporary *decrease* in performance (Dupoux and Green, 1997), while another study shows that a change in compression rate does not affect performance (Golomb *et al.*, 2007). The initial decrease in RT for the time-compressed condition (see Fig. 3) seems to be in line with Golomb *et al.* (2007) that even continuous changes in compression rate did not hinder adaptation to time-compressed speech.

In sum, our results show that listeners adapted to time-compressed speech and natural fast speech and that there was a transfer of learned skills from time-compressed to natural fast sentences, but not the other way around. Adapting to time-compressed speech has been studied extensively in the past decades, and several explanations have been suggested. For instance, adaptation to time-compressed speech has often been described as an attention-weighing process in which listeners shift their attention from task-irrelevant to task-relevant cues (Goldstone, 1998; Golomb *et al.*, 2007; Nosofsky, 1986). Moreover, it has been argued that learning of time-compressed speech is characterized by the recalibration of the boundaries between speech sounds to accommodate the faster speech rate (Golomb *et al.*, 2007). In the discussion below, we attempt to further elucidate the type of cognitive processing underlying adaptation, using Ahissar and Hochstein's (2004) reverse hierarchy theory (RHT), a theory for perceptual learning and transfer (see also Amitay, 2009).

In RHT, perceptual learning is defined as practice-induced improvements in the ability to perform specific perceptual tasks. These improvements involve explicit and extensive practice, for instance, when learning to understand a new language. RHT poses that perceptual learning stems largely from a gradual top-down processing cascade during which first higher and then lower-level task-relevant cues become available. During this process, task-relevant cues are enhanced and task-irrelevant cues are filtered out.

RHT makes explicit predictions about the role of attention and task difficulty on processing level and transfer of learning. With respect to the level of processing, RHT predicts that the cascade from high to low levels of processing is top down and guided by attention as task difficulty increases. When difficulty increases, attention becomes more focused to lower processing levels and lower-level cues become more relevant for task improvement. When applied to our data, this prediction implies that participants relied more on lower-level acoustic cues for conditions that required more attention, i.e., those that were more difficult. It seems plausible that the natural fast condition was the most difficult condition in the experiment as performance was less accurate and slower. Following RHT's prediction, this implies that perceptual learning for the natural fast condition relied more on lower-level acoustic cues than learning of the time-compressed condition. Recall that participants had to process variation resulting from the applied temporal compression while adapting to time-compressed sentences, while for the natural fast sentences they had to adapt to temporal compression *and* to spectral variability. For the natural fast sentences, RHT thus predicts that the higher difficulty of the natural fast sentences condition led them to direct their attention more to lower-level (possibly spectral) acoustic cues than was the case in the time-compressed sentence condition. Further studies are required to address the speculation that spectral and temporal variabilities may be dealt with at different processing levels.

With respect to transfer of learning, RHT predicts that learning at higher processing levels results in more transfer, while learning at lower levels leads to more specificity. RHT also predicts that task difficulty of a preceding task affects learning in the subsequent task. Transfer of learning occurs when an easy condition is followed by a more difficult task, but not when a difficult task is followed by an easier task (Ahissar and Hochstein, 1997; Liu *et al.*, 2008; Pavlovskaya and Hochstein, 2004). Our results comply with this prediction, as we observed task improvement for the natural fast condition (the more difficult task) when it was preceded by the time-compressed condition (the easier task), but not when the natural fast condition preceded the time-compressed condition. Ahissar and Hochstein (2004) suggested that training on easier tasks enables lower-level learning associated with difficult tasks. This suggests for our data that adapting to time-compressed condition, which may involve learning at higher processing levels, improved performance in the natural fast condition by enabling the focus of attention on lower-level cues. As said, learning at lower levels would then lead to more specificity and less transfer from the natural fast to the time-compressed condition.

In conclusion, our results have shown that listeners adapt to extremely fast naturally produced speech. This result is highly relevant because it complements previous research of the learnability of artificially time-compressed speech. Finally, the present results provide one further demonstration of the flexibility of the human speech comprehension system and its ability to adapt on-line to novel variation sources in the speech signal. Our results thus add to a growing body of research on adaptation to natural and artificial variations in the speech signal.

ACKNOWLEDGMENTS

We wish to thank Erik van den Boogert for technical assistance and Matthijs Noordzij for lending his voice. This research was supported by the Netherlands Organization for Research (NWO) under Project Nos. 275-75-003 (P.A.) and 275-75-004 (E.S.).

Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (**2009**). "Familiarity with a regional accent facilitates comprehension of that accent in noise," J. Exp. Psychol. Human **35**, 520–529. AQ: #2

Ahissar, M., and Hochstein, S. (**1997**). "Task difficulty and the specificity of perceptual learning," Nature (London) **387**, 401–406.

Ahissar, M., and Hochstein, S. (**2004**). "The reverse hierarchy of visual perceptual learning," Trends Cogn. Sci. **8**, 457–464.

Amitay, S. (**2009**). "Forward and reserve hierarchies in auditory perceptual learning," Learn. Perception **1**, 59–68. AQ: #3

Baddeley, A. D., Emslie, H., and Nimmo-Smith, I. (**1992**). "The speed and capacity of language processing (SCOLP) test," Bury St Edmunds: Thames Valley Test Company.

Boersma, P., and Weenink, D. (**2003**). "Praat: Doing phonetics by computer," http://www.praat.org (Last viewed 10/28/2006).

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Browman, C., and Goldstein, L. (**1990**). "Tiers in articulatory phonology, with some implications for casual speech," in *Papers in Laboratory Phonology I*, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge), pp. 341–376.

Byrd, D., and Tan, C. C. (**1996**). "Saying consonant clusters quickly," J. Phonetics **24**, 263–282.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cormier, S. M., and Hagman, J. D. (**1987**). *Transfer of Learning: Contemporary Research and Applications*. (Academic, San Diego).

Delhommeau, K., Micheyl, C., and Jouvent, R. (**2005**). "Generalization of frequency discrimination learning across frequencies and ears: Implications for underlying neural mechanisms in humans," J. Assoc. Res. Otolaryngol. **6**, 171–179.

Delhommeau, K., Micheyl, C., Jouvent, R., and Collet, L. (**2002**). "Transfer of learning across durations and across ears in auditory frequency discrimination," Percept. Psychophys. **64**, 426–436.

Dupoux, E., and Green, K. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," Immunopharmacol Immunotoxicol **23**, 914–927.

Ernestus, M., Baayen, H., and Schreuder, R. (**2002**). "The recognition of reduced forms," Brain Lang **81**, 162–173.

Goldstone, R. L. (**1998**). "Perceptual learning," Annu. Rev. Psychol. **49**, 585–612.

Golomb, J., Peelle, J. E., and Wingfield, A. (**2007**). "Effects of stimulus variability and adult aging on adaptation to time-compressed speech," J. Acoust. Soc. Am. **121**, 1701–1708.

Green, K. P., Stevens, K. N., and Kuhl, P. K. (**1994**). "Talker continuity and the use of rate information during phonetic perception," Percept. Psychophys. **55**, 249–260.

Haskell, R. E. (**2001**). *Transfer of Learning: Cognition, Instruction and Reasoning* (Academic, San Diego).

Janse, E. (**2004**). "Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech," Speech Commun. **42**, 155–173.

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

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Liu, E. H., Mercado, E., Church, B. A., and Orduna, I. (**2008**). "The easy-to-hard effect in human (Homo Sapiens) and rat (Rattus Norvegicus) auditory identification," J. Comp. Psychol. **122**, 132–145.

Logan, J. S., Lively, S. E., and Pisoni, D. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Max, L., and Caruso, A. J. (**1997**). "Acoustic measures of temporal intervals across speaking rates: Variability of syllable- and phrase-level relative timing," J. Speech Lang. Hear. Res. **40**, 1097–1110.

May, J., Alcock, K. J., Robinson, L., and Mwita, C. (**2001**). "A computerized test of speed of language comprehension unconfounded by literacy," Appl. Cognit. Psychol. **15**, 433–443.

AQ: #4 McClaskey, C., Pisoni, D., and Carrell, T. (**1983**). "Transfer of learning of a new linguistic contrast in voicing," Percept. Psychophys. **34**, ■.

Miller, J. L., and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Miller, J. L., Aibel, I. L., and Green, K. P. (**1984a**). "On the nature of rate-dependent processing during phonetic perception," Percept. Psychophys. **35**, 5–15.

AQ: #5 Miller, J. L., Grosjean, F., and Lomanto, C. (**1984b**). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implication," Phonetica **41**, 215–225.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.

Nosofsky, R. M. (**1986**). "Attention similarity, and the identification-specific relationship," J. Exp. Psychol. Gen. **115**, 39–57.

Pallier, C., Sebastián-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). "Perceptual adjustment to time-compressed speech: A cross-linguistic study," Mem. Cognit. **26**, 844–851.

Pavlovskaya, M., and Hochstein, S. (**2004**). "Transfer of perceptual learning effects to untrained stimulus dimensions," J. Vis. **4**, 263a–■.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Pinheiro, J., and Bates, D. (**2000**). *Mixed Effects Models in S and S-Plus* (Springer, New York).

Quené, H., and Van den Bergh, H. (**2004**). "On multi-level modeling of data from repeated measures designs: A tutorial," Speech Commun. **43**, 103–121.

Sebastián-Gallés, N., Dupoux, E., Costa, A., and Mehler, J. (**2000**). "Adaptation to time-compressed speech: Phonological determinants," Percept. Psychophys. **62**, 834–842.

Thorndike, E. L., and Woodforth, R. S. (**1901**). "The influence of the improvement in one mental function upon the efficiency of other functions. I.," Psychol. Rev. **8**, 553–656.

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

Wingfield, A., Peelle, J. E., and Grossman, M. (**2003**). "Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults," Aging Neuropsychol. Cogn. **10**, 310–322.

Wouters, J., and Macon, M. W. (**2002**). "Effects of prosodic factors on spectral dynamics. I. Analysis," J. Acoust. Soc. Am. **111**, 417–427.

AUTHOR QUERIES — 018911JAS

#1 Au: PLEASE SUPPLY MISSING ZIP CODES.
#2 Au: The article title in Adank et al. (2009) does not "go" with the rest of the journal citation. Please check: should it be "Comprehension of familiar and unfamiliar native accents under adverse listening conditions"?
#3 Au: Please supply the full journal title, the coden, and/or ISSN for the journal in Amitay (2009) and Pavlovskaya and Hochstein (2004).
#4 Au: PLEASE SUPPLY PAGE RANGE IN McClaskey et al. (1983). AND Pavlovskaya and Hochstein (2004).
#5 Au: Please provide text citation for Miller et al. (1984b) or delete.
#6 Author: Please provide a caption for Table I.