Oscillatory phase shapes syllable perception

Sanne ten Oever[1] and Alexander T. Sack

Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6229 EV Maastricht, The Netherlands

Edited by Randolph Blake, Vanderbilt University, Nashville, TN, and approved November 9, 2015 (received for review September 2, 2015)

The role of oscillatory phase for perceptual and cognitive processes is being increasingly acknowledged. To date, little is known about the direct role of phase in categorical perception. Here we show in two separate experiments that the identification of ambiguous syllables that can either be perceived as /da/ or /ga/ is biased by the underlying oscillatory phase as measured with EEG and sensory entrainment to rhythmic stimuli. The measured phase difference in which perception is biased toward /da/ or /ga/ exactly matched the different temporal onset delays in natural audiovisual speech between mouth movements and speech sounds, which last 80 ms longer for /ga/ than for /da/. These results indicate the functional relationship between prestimulus phase and syllable identification, and signify that the origin of this phase relationship could lie in exposure and subsequent learning of unique audiovisual temporal onset differences.

oscillations | phase | audiovisual | speech | temporal processing

In spoken language, visual mouth movements naturally precede the production of any speech sound, and therefore serve as a temporal prediction and detection cue for identifying spoken language (1) (but also see ref. 2). Different syllables are characterized by unique visual-to-auditory temporal asynchronies (3, 4). For example, /ga/ has an 80-ms longer delay than /da/, and this difference aids categorical perception of these syllables (4). We propose that neuronal oscillations might carry the information to dissociate these syllables based on temporal differences. Multiple authors have proposed (5–7)—and it has been demonstrated empirically (7–9)—that at the onset of visual mouth movements, ongoing oscillations in auditory cortex align (see refs. 10–12 for nonspeech phase reset), providing a temporal reference frame for the auditory processing of subsequent speech sounds. Consequently, auditory signals fall on different phases of the aligned oscillation depending on the unique visual-to-auditory temporal asynchrony, resulting in a consistent relationship between syllable identity and oscillatory phase.

We hypothesized that this consistent "phase–syllable" relationship results in ongoing oscillatory phase biasing syllable perception. More specifically, the phase at which syllable perception is mostly biased should be proportional to the visual-to-auditory temporal asynchrony found in natural speech. A naturally occurring /ga/ has an 80-ms longer visual-to-auditory onset difference than a naturally occurring /da/ (4). Consequently, the phase difference between perception bias toward /da/ and /ga/ should match 80 ms, which can only be established with an oscillation with a period greater than 80 ms, that is, any oscillation under 12.5 Hz. The apparent relevant oscillation range is therefore theta, with periods ranging between 111 and 250 ms (4–9 Hz). This oscillation range has already been proposed as a candidate to encode information, and seems specifically important for speech perception (13, 14).

To test this hypothesis of oscillatory phase biasing auditory syllable perception in the absence of visual signals, we presented ambiguous auditory syllables that could be interpreted as /da/ or /ga/ while recording EEG. In a second experiment, we used sensory entrainment (thereby externally enforcing oscillatory patterns) to demonstrate that entrained phase indeed determines whether participants identify the presented ambiguous syllable as being either /da/ or /ga/.

Results

Experiment 1.

Psychometric curves. First, we created nine morphs between a /da/ and a /ga/ by shifting the third formant frequency of a recorded /da/ from around 2,600–3,000 Hz (Fig. 1*A*). We determined the individual threshold at which participants would identify a morphed stimulus 50% as /da/ and 50% as /ga/ by repeatedly presenting the nine different morphs, and participants had to indicate their percept (see *SI Materials and Methods* for details). Indeed, 18 out of 20 participants were sensitive to the manipulation of the morphed stimulus, and psychometric curves could be fitted reliably (Fig. 1*B*; average explained variance of the fit was 92.7%, SD of 0.03). The other two participants were excluded from further analyses.

Consistency of phase differences. We used the individually determined most ambiguous stimuli to investigate whether ongoing theta phase before stimulus presentation influenced the identification of the syllable. Therefore, we presented both the unambiguous /da/ (stimulus 1) and /ga/ (stimulus 9) and the ambiguous stimulus while recording EEG. Data were epoched −3 to 3 s around syllable onset. To ensure that poststimulus effects did not temporally smear back to the prestimulus interval (e.g., 15), we padded all data points after zero with the amplitude value at zero. For every participant, we extracted the average phase for each of the syllable types for the −0.3- to 0.2-s interval. There were four syllable types: the /da/ and /ga/ of the unambiguous sounds and the ambiguous sound either perceived as /da/ or /ga/. Then, we determined the phase difference between /da/ and /ga/ for both the unambiguous and ambiguous conditions. In the ambiguous condition, prestimulus phase is hypothesized to bias syllable perception, and this should be reflected in a consistent phase difference between the perceived /da/ and /ga/. In the unambiguous condition in the prestimulus phase, time windows should mostly reflect random fluctuations, because participants are unaware of the identity and arrival time of the upcoming syllable and generally identified stimulus 1 as /da/ and stimulus 9 as /ga/, resulting in a low consistency of the phase

Significance

The environment is full of temporal information that links specific auditory and visual representations to each other. Especially in speech, this is used to guide perception. The current paper shows that syllables with varying visual-to-auditory delays get preferably processed at different oscillatory phases. This mechanism facilitates the separation of different representations based on consistent temporal patterns in the environment and provides a way to categorize and memorize information, thereby optimizing a wide variety of perceptual processes.

Author contributions: S.t.O. designed research; S.t.O. performed research; S.t.O. analyzed data; and S.t.O. and A.T.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. Email: sanne.tenoever@maastrichtuniversity.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1517519112/-/DCSupplemental.

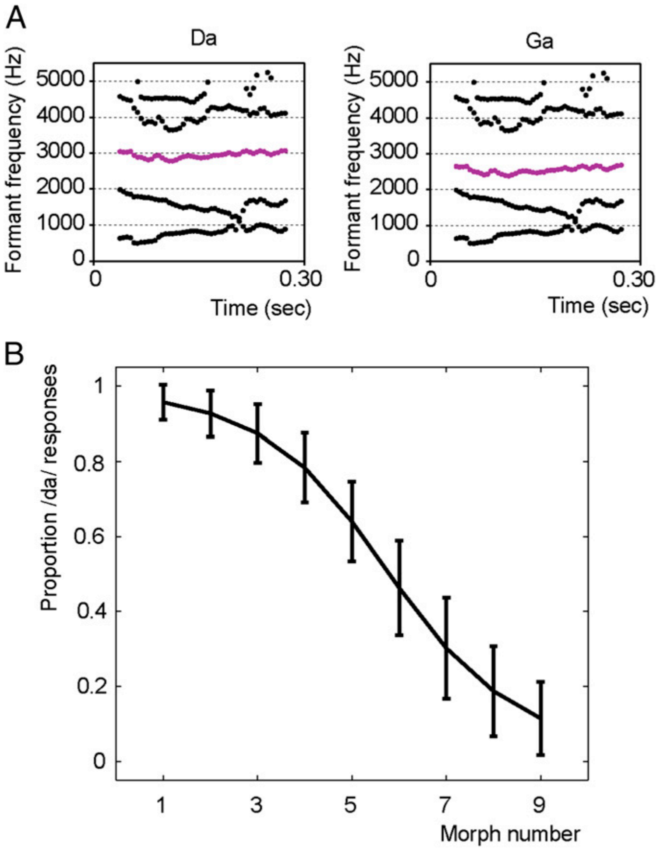

Fig. 1. Results from morphed /daga/ stimuli. (*A*) Stimulus properties of the used /da/ and /ga/ stimuli. Only the third formant differs between the two stimuli (purple lines). (*B*) Average proportion of /da/ responses for the 18 participants in experiment 1. Error bars reflect the SEM.

difference. Note, however, that, in principle, phase differences are possible in this condition, because we did exclude trials in which participants identified the unambiguous syllables as the syllable at the other side of the morphed spectrum. The mean resultant vector lengths (MRVLs) of the phase difference between /da/ and /ga/ were calculated, and Monte Carlo simulations with a cluster-based correction for multiple comparisons were used for statistical testing. A higher MRVL indicates a higher phase concentration of the difference. We found that the ambiguous sounds had a significantly higher MRVL before sound onset (−0.25 to −0.1 ms) around 6 Hz (cluster statistics 19.821, P = 0.006; Fig. 2 *A* and *B*). When repeating the analysis including a wider frequency spectrum (1–40 Hz), the same effect was present (cluster statistics 18.164, P = 0.030), showing the specificity of the effect for theta. Because any phase estimation requires integration of data over time, the significant data appear distant from the onset of the syllable. For example, the 6-Hz phase angle is calculated using a window of 700 ms (to ensure the inclusion of multiple cycles of the theta oscillation). The closer the center of the estimation is to an abrupt change in the data (such as a stimulus or the data padding to zero), the more the estimation is negatively influenced by the “postchange data” (e.g., 15).

Eighty-millisecond phase differences. A second hypothesis was that the phase difference of the ambiguous stimuli judged as /da/ vs. /ga/ would match 80 ms, consistent with the visual-to-auditory onset difference between /da/ and /ga/ found in natural speech (4). Therefore, we took all of the significant time and frequency points in the first analysis and tested whether the phase difference of all participants was centered around 80 ms (the blue line in Fig. 2*B* corresponds to an 80-ms difference). This is typically done with the *V* test, which examines the nonuniformity of circular data centered around a known specific mean direction. We found that the ambiguous phase differences indeed centered around 80 ms for almost all tested data points, whereas for the unambiguous sounds no such phase concentration was present (Fig. 2*C*).

From Fig. 2*B* it is evident that there is a consistent phase difference across participants between /da/ and /ga/ for the ambiguous sounds. When looking at the consistency of the phases of the individual syllables /da/ and /ga/ this consistency drops (compare Fig. 2*B* with Fig. S1*B*). Statistical testing confirmed that the /da/ and /ga/ phases seemed distributed randomly (Fig. S1*C*). At this point, we cannot differentiate whether this effect occurs due to volume conduction of the EEG or individual latency differences for syllable processing (see also ref. 12). When repeating this analysis for each participant, we did find a significant (uncorrected) consistency for multiple participants and a significant different phase between /da/ and /ga/ (Fig. S2; for only two participants this effect survived correction for multiple comparisons).

The current reported effects could not be explained by any eye movements (no significant differences between conditions) or any artifacts due to data padding (Fig. S3).

Experiment 2. To investigate whether neuronal entrainment results in oscillatory identification patterns, we experimentally induced theta phase alignment using sensory entrainment (16–18) in 12 different participants. In this experiment, auditory stimuli of broadband noise (white noise band pass-filtered between 2.5 and 3.1 kHz, 50-ms length) were repeatedly presented (presumably entraining underlying oscillations at the presentation rate), after which ambiguous sounds were presented at different stimulus onset asynchronies (SOAs; 12 different SOAs fitting exactly two cycles). If ongoing phase is important for syllable identification, the time course of identification should oscillate at the presentation rate. Indeed, the time course of identification showed a pattern varying at the presentation rate of 6.25 Hz (Fig. 3*A*). To test the significance of this effect, we calculated the relevance value (19). This value is calculated by (*i*) fitting a sinus to the data and (*ii*) multiplying the explained variance of the fit by the variance of the predicted values. In this way, the relevance statistic gives less weight to models that have a fit with a flat line. Thereafter, we performed bootstrapping on the obtained relevance values (of the average curve) to show that of the 10,000 fitted bootstraps only 2.83% had a more extreme relevance value (Fig. 3*B*), suggesting that, indeed, syllable identity depends on theta phase.

Three control experiments were performed. In the first two experiments, the frequency specificity of the effect was investigated by changing the presentation rates of the entrainment train to 1 and 10 Hz. In a third experiment, we wanted to rule out the possibility that the effect already occurs at a lower perceptual level instead of the syllable identification level. Therefore, we band pass-filtered the syllables between 2.5 and 3.1 Hz, maintaining the formant frequency at which the two syllables differ but distorting syllable perception. Participants had to indicate whether they felt the sound was of high or low frequency (this experiment will from now on be called “frequency control”). As a reference for what was considered a high or low frequency, the band pass-filtered stimulus numbers 1 and 9 were both presented at random order at the beginning of the trial.

Results show that for both the 1-Hz and the frequency control, no sinus could be fitted reliably (Fig. 3 *B* and *C*; P = 0.80 and P = 0.69, respectively). In contrast, for 10 Hz, a sinus could be reliably fitted (P = 0.011). For all three presentation frequencies there was entrainment at the expected frequency (Fig. 3*D*).

Discussion

In the current study, we investigated whether ongoing oscillatory phase biases syllable identification. We presented ambiguous

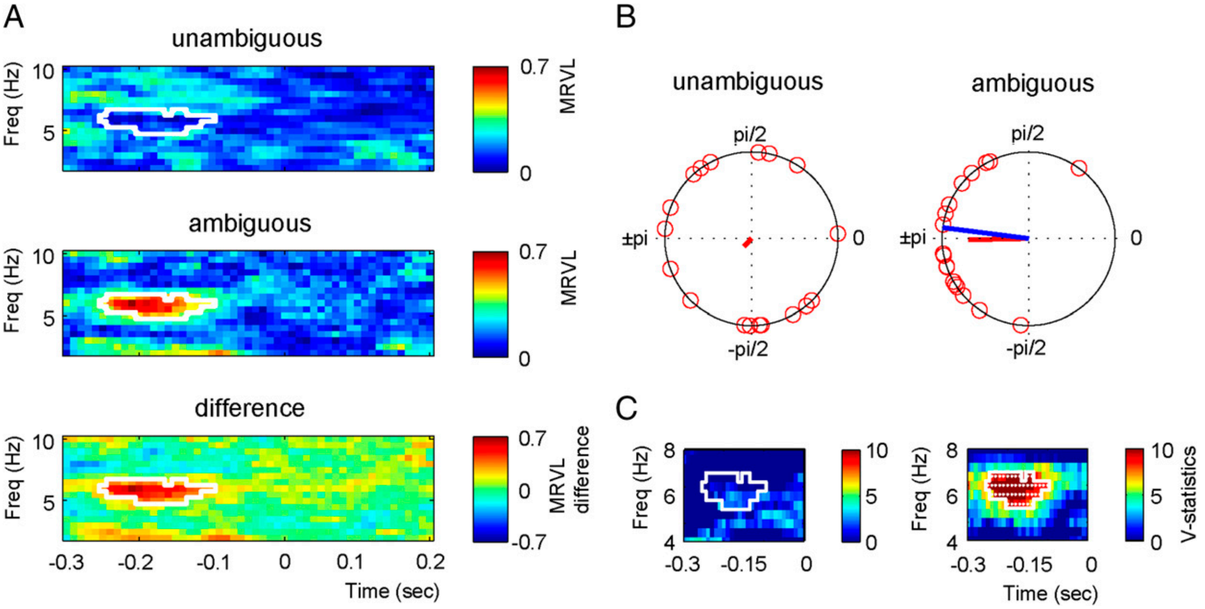

Fig. 2. Prestimulus phase differences. (*A*) The mean resultant vector length across participants for the phase difference between /da/ and /ga/ for unambiguous sounds and for the phase difference between perceived /da/ and /ga/ for ambiguous sounds. The white outlines indicate the region of significant differences. (*B*) Phase differences of individual participants at 6 Hz at −0.18 s for unambiguous and ambiguous sounds. The blue line indicates the 80-ms expected difference. The red lines indicate the strength of the MRVL. (*C*) *V* statistics testing whether the phase differences are significantly nonuniformly distributed around 80 ms for all significant points in the MRVL analysis. The white outlines indicate at which time and frequency point the analysis was performed (note the difference in the *x* and *y* axes between *A* and *C*). White dots indicate significance.

auditory stimuli while recording EEG and revealed a systematic phase difference before auditory onset between the perceived /da/ and /ga/ at theta frequency. This phase discrepancy corresponded to the 80-ms difference between the onset delays of the speech

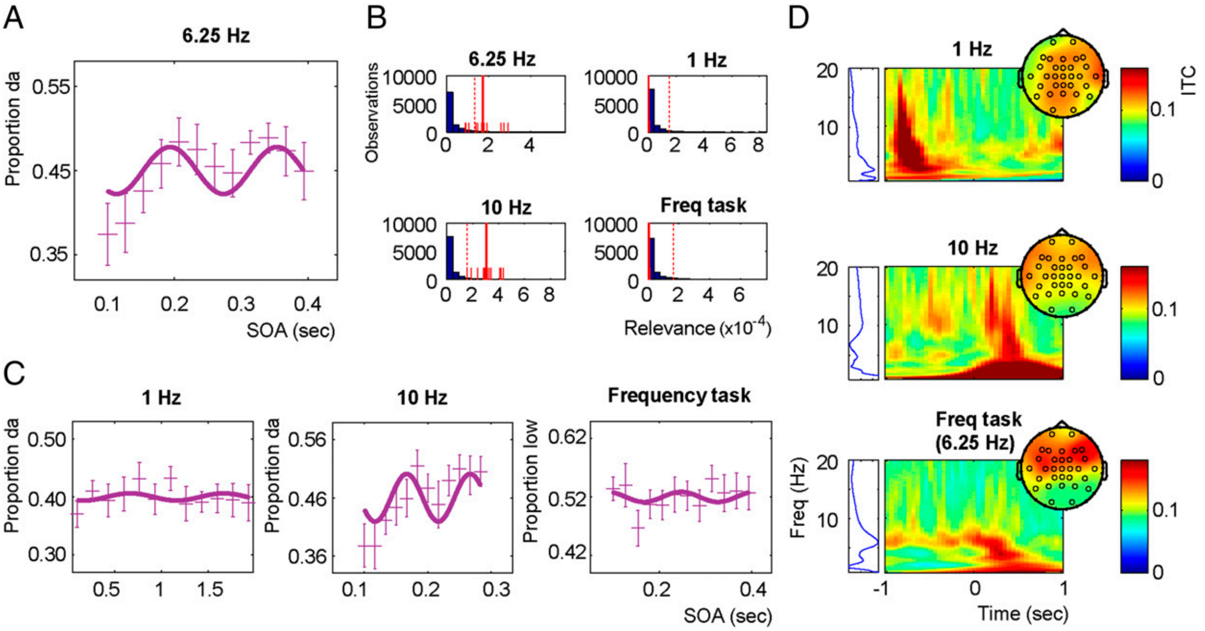

Fig. 3. Results from experiment 2. (*A*) Grand average proportion of /da/ of all of the participants, with the respective error bars reflecting the within-subject SEM (plusses; vertical extension reflects the error bars) and the fitted 6.25-Hz sinus (solid line). (*B*) Bootstrap histograms for the relevance statistics for all four conditions. The long solid and dotted red lines represent the relevance value of that dataset and the 95 percentile of all bootstrapped values, respectively. The short solid lines indicate the 12 relevance values when iteratively taking out one participant. The blue bars represent the individual relevance values of all the different bootstraps. (*C*) The grand average of all participants, with the respective error bars reflecting the within-subject SEM (plusses; vertical extension reflects the error bars) for the three different control conditions used in the experiment and their respective best-fitted sinus (solid line). (*D*) Intertrial coherence (ITC) plots for all three entrainment frequencies. Zero indicates entrainment offset. (*Left*) The ITC averaged in the −0.5 to 0 interval (ITC range 0.08–0.12). All of the conditions show a peak at the respective entrainment frequency. However, for 1 Hz, an evoked response of the last entrainment stimulus is present (around −0.8 s). For 10 Hz, and to a lesser extent for 6.25 Hz, evoked responses to the target stimuli are present poststimulus (around 0–1 s). This effect only arises in these frequencies, because the interval target presented is much narrower than for 1 Hz.

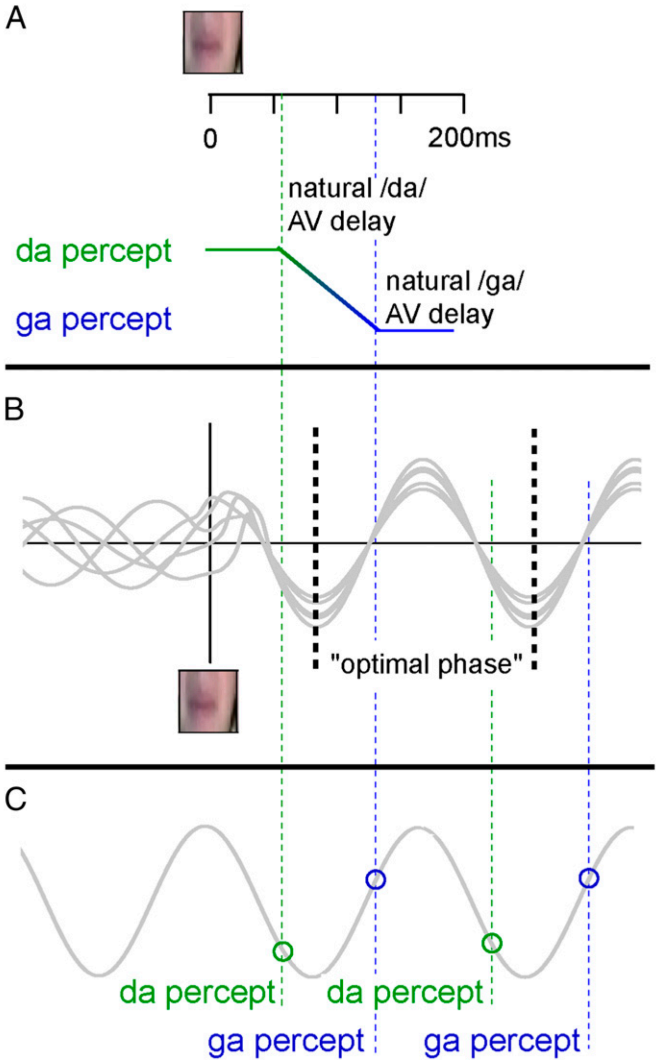

Fig. 4. Proposed mechanism for theta phase sensitization. (*A*) Dependent on the natural visual-to-auditory (AV) delay, voiced-stop consonants are identified as a /da/ or a /ga/ after presenting the same visual stimulus (4). (*B*) When visual speech is presented, ongoing theta oscillations synchronize, creating an optimal phase (black dotted line) at which stimuli are best-processed. The phase at which a /da/ or a /ga/ in natural situations is presented is different (green and blue lines, respectively), caused by the difference in visual-to-auditory delay. (*C*) Syllable perception is biased at phases at which /da/ and /ga/ are systematically presented in audiovisual settings even when visual input is absent.

sounds /da/ or /ga/ with respect to the onset of the corresponding mouth movements found in natural speech (4). Moreover, we showed that syllable identification depends on the underlying oscillatory phase induced by entrainment to a 6.25- or 10-Hz presented stimulus train of broadband noise. These results reveal the relevance of phase coding for language perception and provide a flexible mechanism for statistical learning of onset differences and possibly for the encoding of other temporal information for optimizing perception.

Audiovisual Learning Results in Phase Coding. The human brain is remarkably capable of associating events that repeatedly occur together (20, 21), representing an efficient neural coding mechanism for guiding our interpretation of the environment. Specifically, when two events tend to occur together, they will enhance the neural connections between each other, consequently increasing the detection sensitivity of one event in case the associated event is present (22). We propose that this could also work for temporal associations. In a previous study, we showed that the onset between mouth movements and auditory speech signals differs between syllables, and that this influences syllable identification (4). For example, a naturally occurring /ga/ has an 80-ms larger visual-to-auditory onset difference than a naturally occurring /da/ (Fig. 4*A*) (4). Recent theories propose that visual cues benefit auditory speech processing by aligning ongoing oscillations in auditory cortex such that the "optimal" high excitable period coincides with the time point at which auditory stimuli are expected to arrive, thereby optimizing their processing (Fig. 4*B*) (8, 10, 23). If this indeed occurs, different syllables should be consistently presented at different phases of the reset oscillation (green and blue lines in Fig. 4*B*). A similar mechanism has also been proposed by Peelle and Davis (14). Because humans (or rather our brains) likely (implicitly) learn this consistent association between phase and syllable identity, one could hypothesize that neuronal populations coding for different syllables may begin to prefer specific phases, biasing syllable perception at corresponding phases even when visual input is absent (Fig. 4*C*). The current data indeed support this notion, as we show that the phase difference between /da/ and /ga/ fits 80 ms. The exact cortical origin of this effect cannot be unraveled with the current data, but we would expect to find these effects in auditory cortex.

Generalization of This Mechanism. Temporal information is not only present in (audiovisual) speech. Therefore, any consistent temporal relationship between two stimuli could be coded in a similar vein as demonstrated here. For example, the proposed mechanism should also generalize to auditory-only settings, because any temporal differences caused by articulatory processes should also influence the timing of individual syllables within a word; for example, the second syllable in "ba*ga*" should arrive at a later time point as "ba*da*." It is an open question how these types of mechanisms generalize to situations in which speech is faster or slower. However, it is conceivable that when speaking faster the visual-to-auditory onset differences between /da/ and /ga/ also reduce, thereby also changing their expected phase difference. It has already been shown that cross-modal mechanisms rapidly update changing temporal statistics in the environment (24), by for example changing the oscillatory phase relationship between visual and auditory regions (25).

Our results show that during 10-Hz entrainment an oscillatory pattern of syllable identification is present. This frequency is slightly higher than what is generally considered theta. This likely reflects that the brain flexibly adapts to the changing environment, for example when facing a person who speaks very fast. Thus, although under "normal" circumstances the effect seems constrained to theta (as shown in experiment 1), altering the brain state by entraining to higher frequencies still induces the effect and shows the flexibility of this mechanism.

Excitability Versus Phase Coding. Much research has focused on the role of oscillations in systematically increasing and decreasing the excitability levels of neuronal populations (23, 26, 27). In this line of reasoning, speech processing is enhanced by aligning the most excitable phase of an oscillation to the incoming speech signal (5, 6). Intuitively, our results seem in contrast to this idea, as it appears that neuronal populations coding for separate syllables have phase-specific responses. However, it could also be considered possible that one neuronal population biases identification in the direction of one syllable, this bias succeeding when excited and failing when suppressed. This interpretation is less likely, considering that the exact phases at which syllable identification was biased varied across participants. Therefore, the phase at which identification is biased toward one syllable does not always fall on the most excitable point of the oscillation for each participant (unless the phases of the measured EEG signal are not comparable across participants). Considering that there are individual differences in the lag between stimulus presentation and brain response (e.g., 18), it would also follow that the phase at which syllable identification is biased does not match across participants. However, more research is needed to irrefutably demonstrate that different neuronal populations code information preferably at a specific oscillatory phase (28).

Conclusion

Temporal associations are omnipresent in our environment, and it seems highly unlikely that these data are ignored by our brain when information has to be ordered and categorized. The current study has demonstrated that oscillatory phase shapes syllable perception and that this phase difference matches temporal statistics in the environment. To determine whether this type of phase sensitization is a common neural mechanism, it is necessary to investigate other types of temporal statistics, especially because it could provide a mechanism for separating different representations (26, 29, 30) and offer an efficient way of coding time differences (31). Future research needs to investigate whether also other properties are encoded in phase, revealing the full potential of this type of phase coding scheme.

Materials and Methods

In total, 40 participants took part in our study (20 per experiment). All participants gave written informed consent. The study was approved by the local ethical committee at the Faculty of Psychology and Neuroscience at Maastricht University. Detailed methods are described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Giancarlo Valente for useful suggestions for the analysis. Kirsten Petras and Helen Lückmann helped improve the final manuscript. This study was supported by a grant from the Dutch Organization for Scientific Research (NWO; 406-11-068).

1. Campbell R (2008) The processing of audio-visual speech: Empirical and neural bases. *Philos Trans R Soc Lond B Biol Sci* 363(1493):1001–1010.
2. Schwartz J-L, Savariaux C (2014) No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Comput Biol* 10(7):e1003743.
3. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5(7):e1000436.
4. ten Oever S, Sack AT, Wheat KL, Bien N, van Atteveldt N (2013) Audio-visual onset differences are used to determine syllable identity for ambiguous audio-visual stimulus pairs. *Front Psychol* 4:331.
5. Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12(3):106–113.
6. Peelle JE, Sommers MS (2015) Prediction and constraint in audiovisual speech perception. *Cortex* 68:169–181.
7. Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8(8): e1000445.
8. van Atteveldt N, Murray MM, Thut G, Schroeder CE (2014) Multisensory integration: Flexible use of general operations. *Neuron* 81(6):1240–1253.
9. Perrodin C, Kayser C, Logothetis NK, Petkov CI (2015) Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proc Natl Acad Sci USA* 112(1):273–278.
10. Mercier MR, et al. (2015) Neuro-oscillatory phase alignment drives speeded multisensory response times: An electro-corticographic investigation. *J Neurosci* 35(22): 8546–8557.
11. Besle J, et al. (2011) Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci* 31(9):3176–3185.
12. Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53(2):279–292.
13. Kayser C, Ince RA, Panzeri S (2012) Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices. *PLoS Comput Biol* 8(10):e1002717.
14. Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. *Front Psychol* 3:320.
15. Zoefel B, Heil P (2013) Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front Psychol* 4:262.
16. de Graaf TA, et al. (2013) Alpha-band rhythms in visual task performance: Phase-locking by rhythmic sensory stimulation. *PLoS One* 8(3):e60035.
17. Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320(5872):110–113.
18. Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc Natl Acad Sci USA* 109(49): 20095–20100.
19. Fiebelkorn IC, et al. (2011) Ready, set, reset: Stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *J Neurosci* 31(27):9971–9981.
20. Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13(9):403–409.
21. Fiser J, Aslin RN (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol Sci* 12(6):499–504.
22. Hebb DO (1949) *The organization of behavior: A neuropsychological theory* (John Wiley & Sons, New York).
23. Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32(1):9–18.
24. Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7(7):773–778.
25. Kösem A, Gramfort A, van Wassenhove V (2014) Encoding of event timing in the phase of neural oscillations. *Neuroimage* 92:274–284.
26. Jensen O, Gips B, Bergmann TO, Bonnefond M (2014) Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci* 37(7):357–369.
27. Mathewson KE, Fabiani M, Gratton G, Beck DM, Lleras A (2010) Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition* 115(1):186–191.
28. Watrous AJ, Fell J, Ekstrom AD, Axmacher N (2015) More than spikes: Common oscillatory mechanisms for content specific neural representations during perception and memory. *Curr Opin Neurobiol* 31:33–39.
29. Fell J, Axmacher N (2011) The role of phase synchronization in memory processes. *Nat Rev Neurosci* 12(2):105–118.
30. Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61(4):597–608.
31. Chakravarthi R, Vanrullen R (2012) Conscious updating is a rhythmic process. *Proc Natl Acad Sci USA* 109(26):10599–10604.
32. Boersma P, Weenink D (2013) Praat: Doing Phonectics by Computer (University of Amsterdam, Amsterdam), Version 5.3.56.
33. Bertelson P, Vroomen J, De Gelder B (2003) Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychol Sci* 14(6):592–597.
34. Zychaluk K, Foster DH (2009) Model-free estimation of the psychometric function. *Atten Percept Psychophys* 71(6):1414–1425.
35. Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
36. Berens P (2009) CircStat: A MATLAB toolbox for circular statistics. *J Stat Softw* 31(10): 1–21.
37. Gómez-Herrero G, et al. (2006) Automatic removal of ocular artifacts in the EEG without an EOG reference channel. *Proceedings of the 7th Nordic Signal Processing Symposium* (IEEE, Rejkjavik), pp 130–133.
38. Zar JH (1998) *Biostatistical Analysis* (Prentice Hall, Englewood Cliffs, NJ), 4th Ed.
39. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4):1165–1188.
40. Mewhort DJ, Kelly M, Johns BT (2009) Randomization tests and the unequal-N/unequal-variance problem. *Behav Res Methods* 41(3):664–667.
41. Hari R, Hämäläinen M, Joutsiniemi SL (1989) Neuromagnetic steady-state responses to auditory stimuli. *J Acoust Soc Am* 86(3):1033–1039.
42. Rees A, Green GG, Kay RH (1986) Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res* 23(2):123–133.
43. Lakatos P, et al. (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77(4):750–761.