1. In general, it is difficult to support a theory about articulatory coordination using acoustic data, since theories of articulatory coordination hinge on the relative timing of articulatory events (gesture activation onsets in the case of Nam et al. 2010). In the current case, the details of the proposed account are not sufficient to evaluate it with respect to the predictions that it makes for the data presented in the paper, or for other known facts about speech timing. Particularly relevant are facts and theories about polysyllabic shortening, which are not discussed in the paper in spite of the monosyllable vs. disyllable manipulation, as well as facts and theories about differences in movement timing for voiced vs. voiceless consonants in other positions-in-syllable, e.g. syllable onset position. Note that a plausible alternative account of the attenuation of the voicing effect in disyllables could come from a higher weighting of factors such as the number of syllables in a word and/or position in the word compared to the voicing factor.
2. In particular, there are several issues about the account that require clarification before the proposal can be evaluated:
    a. Assumptions about syllabification –to what extent does the proposed gestural account hinge on an assumption that C2 is syllable-initial, as suggested by your proposal that it coordinates in-phase with a following vowel? This is a controversial point, cf. phonetic and allophonic behaviour suggesting that consonants in this position are either ambisyllabic or syllable-final.
    b. What are the assumptions about gestures in your proposed gestural account? What are gestures exactly? Are they as defined in Browman & Goldstein (1992), Saltzman et al. (2008)?
    c. Your account proposes a difference in gestural stiffness between voiced and voiceless consonants. Is this difference part of phonological representation? If so, how would you account for the presumed lack of velocity differences in word-initial position (cf. Ostry & Munhall 1985)? If not, how does this change get implemented?
    d. What about gestural activation? How are gestural activation intervals determined?
3. Note that the "voicing effect" is not found in all varieties of English. Scottish English does not show the same effect as other varieties.

References

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica, 49*(3-4), 155-180.

Nam, H., Goldstein, L., & Saltzman, E. (2010). Self-organization of syllable structure: a coupled oscillator model. In François Pellegrino, E. Marisco, & I. Chitoran (Eds.), *Approaches to phonological complexity*. Berlin, New York: Mouton de Gruyter.

Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America, 77*, 640-648.

Saltzman, E., Nam, H., Krivokapić, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira, & C. Reis (Eds.), *Proceedings of the Speech Prosody 2008 Conference* (pp. 175-184). Campinas, Brazil.