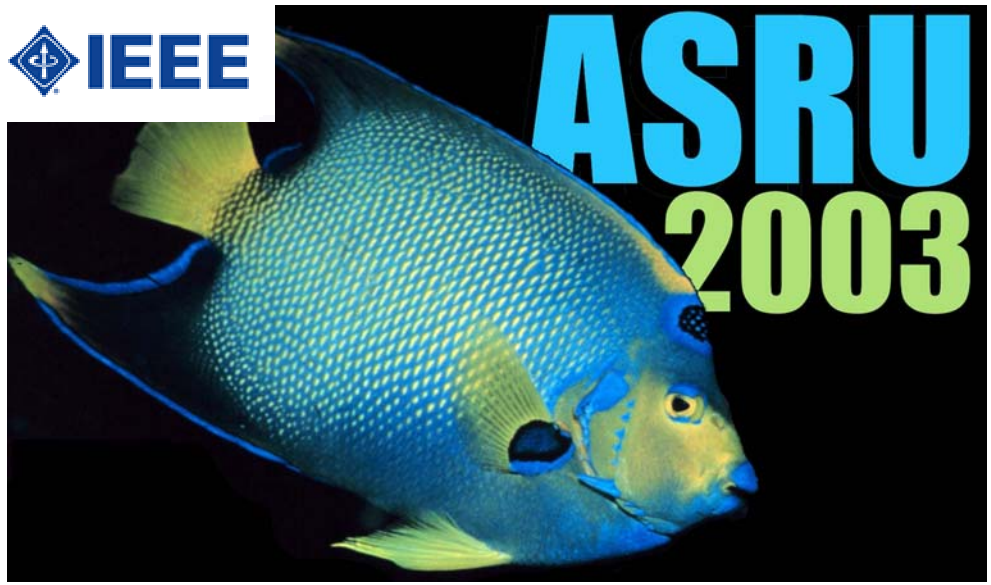


Speculating on the Future for Automatic Speech Recognition

A Survey of Attendees

by

Roger K Moore

elsnet

aurix

THANK YOU !

"It is hard to predict …"

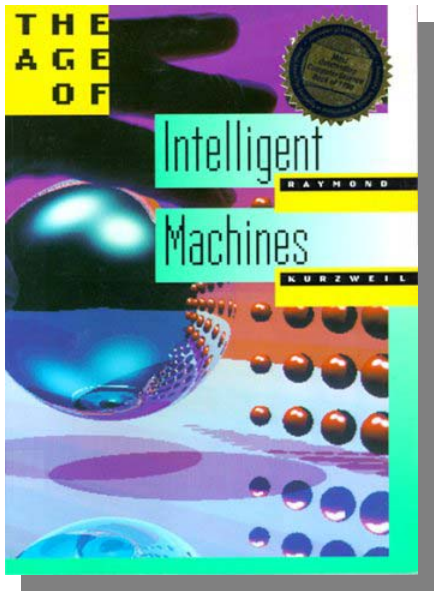
"… especially the future."

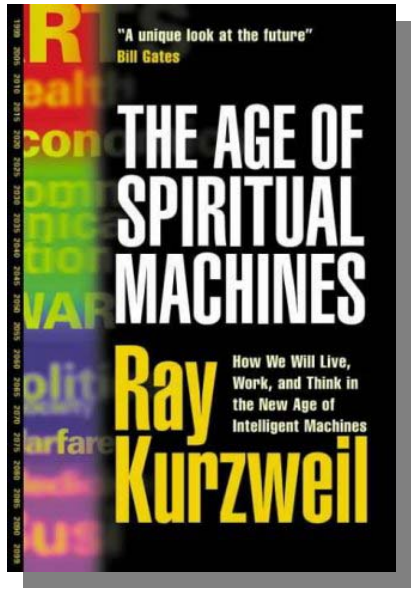
Niels Bohr, 1922

The Survey(s)

- 12 of the 20 statements were exactly the same as those posed to the participants of ASRU'97 six years ago
- A couple were suggested by the ASRU'04 Technical Committee
- The rest were taken from Ray Kurzweil's books …

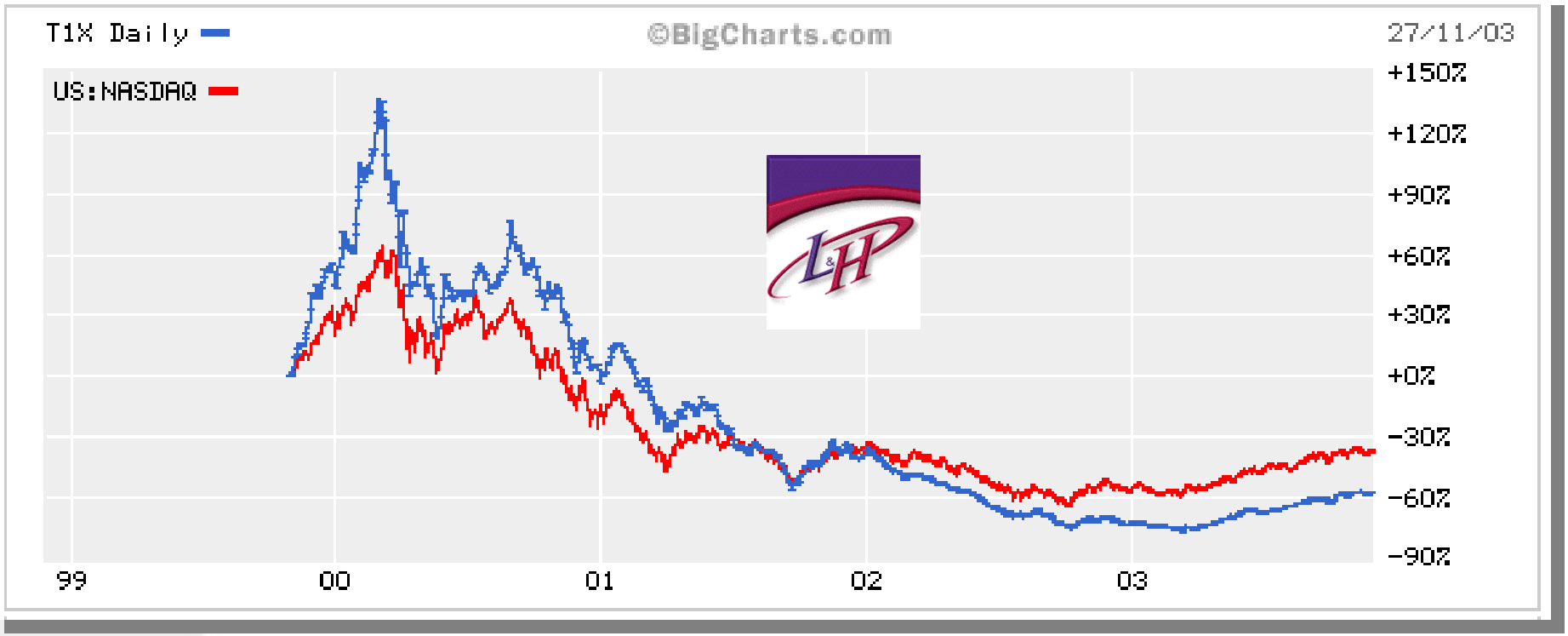
Predictions from Ray Kurzweil





"A PC will have the computational power of the human brain by 2019, and will be equivalent to 1000 human brains by 2029."

1997 - 2003



PC PRODUCTS
DRAGON SYSTEMS

Some Overall Statistics

attendees:	222
forms returned:	47%
overall mean:	2055
“never”s:	24%
named responses:	4

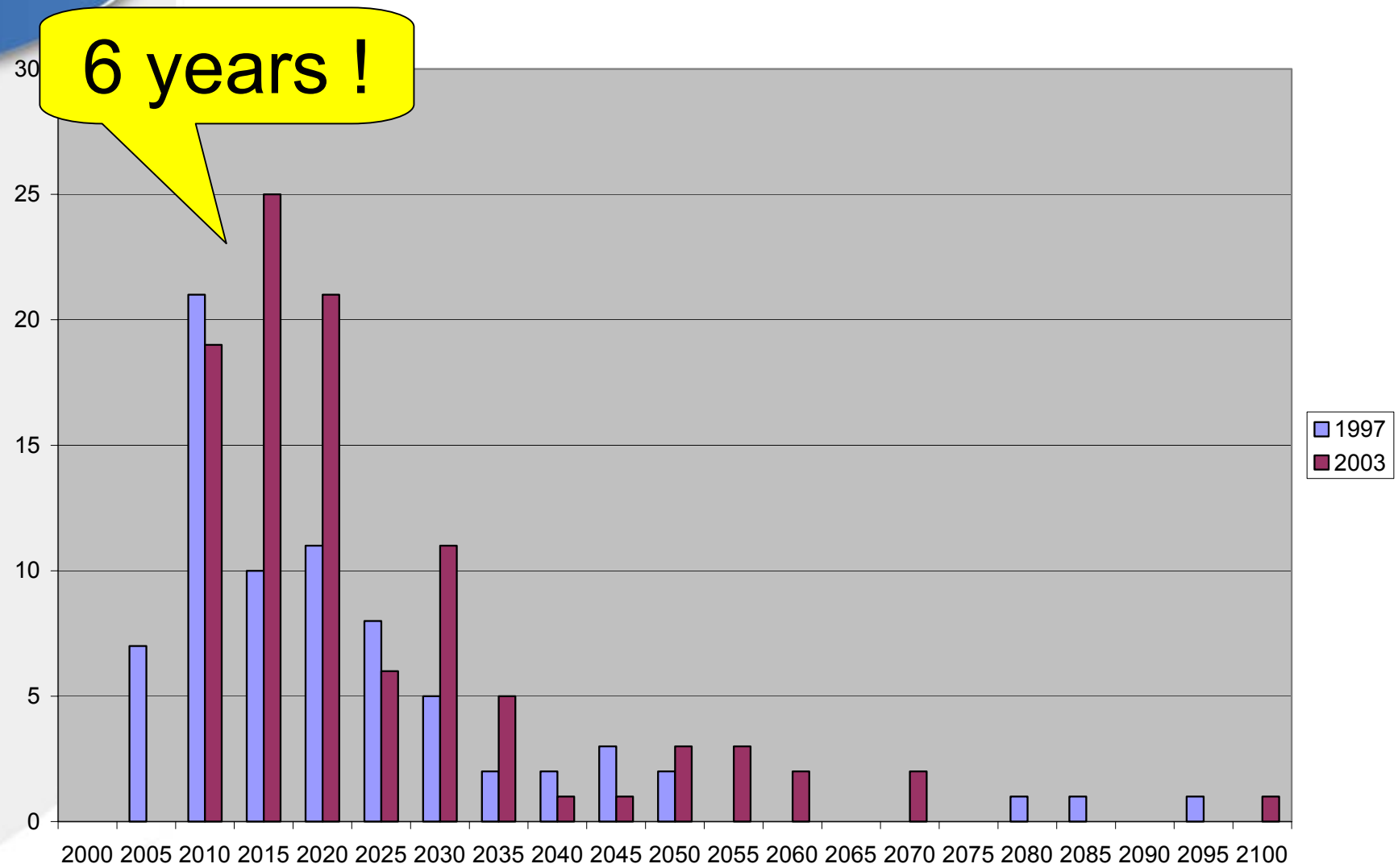
Some Overall Statistics

	2003	(1997)
attendees:	222	(180)
forms returned:	47%	(45%)
overall mean:	2055	(2056)
“never”s:	24%	(17%)
named responses:	4	18

Some Overall Statistics

	2003	(1997)
attendees:	222	(180)
forms returned:	47%	(45%)
overall mean:	2055	(2056)
"never"s:	24%	(17%)
named responses:	4	18
"2020"s:	10%	(7%)

The ‘Church Effect’



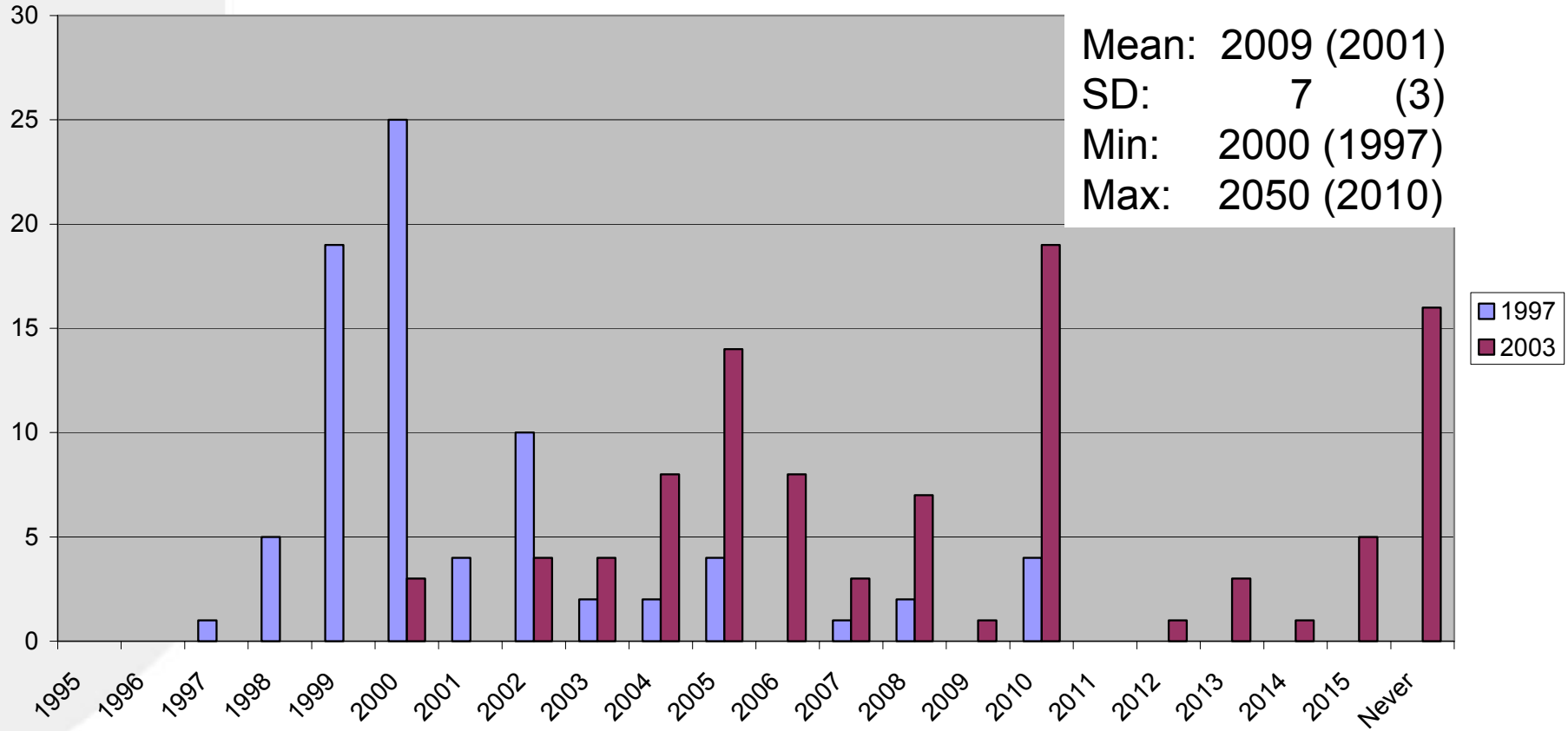
1. More than 50% of new PCs have dictation on them, either at purchase or shortly after.

"won't be used"

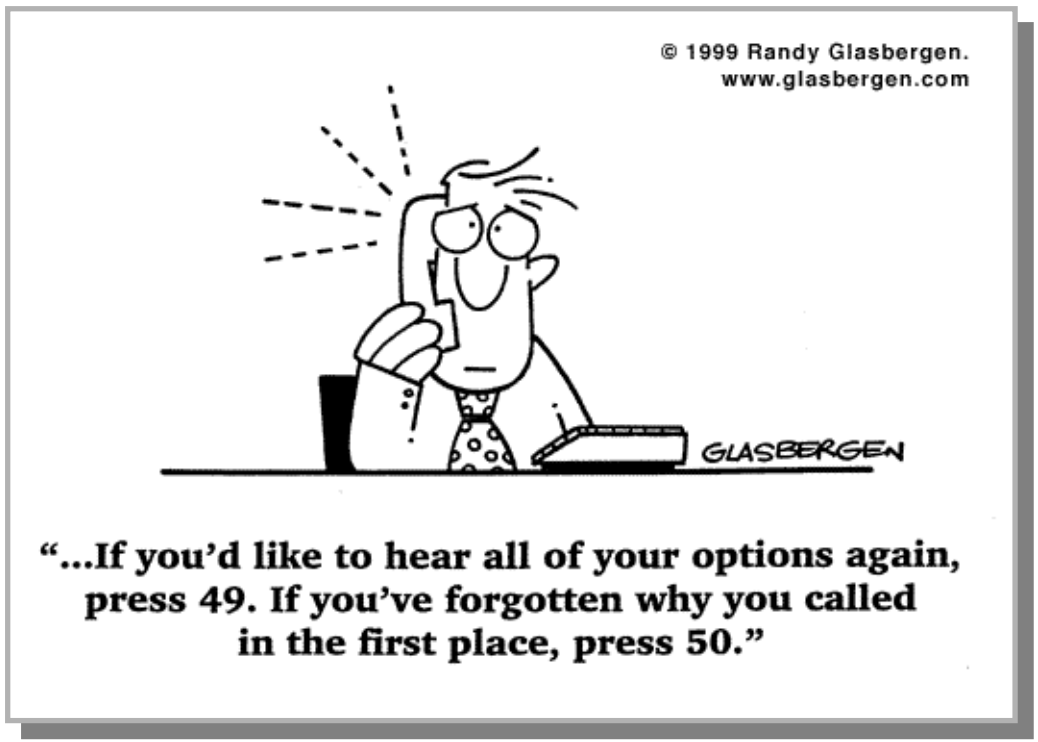
"already comes with Office XP"

"now … but not used"

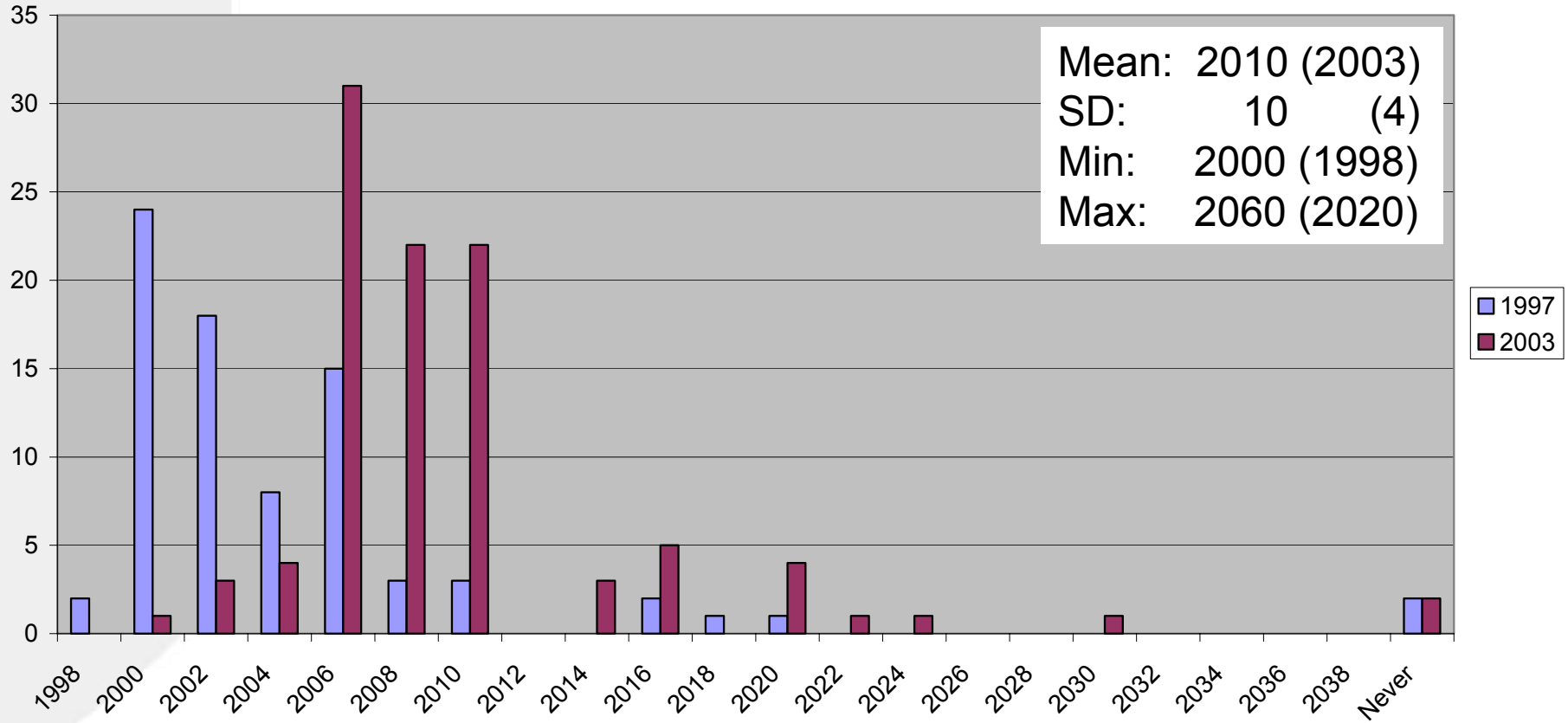
1. More than 50% of new PCs have dictation on them, either at purchase or shortly after.



2. Most telephone Interactive Voice Response (IVR) systems accept speech input.

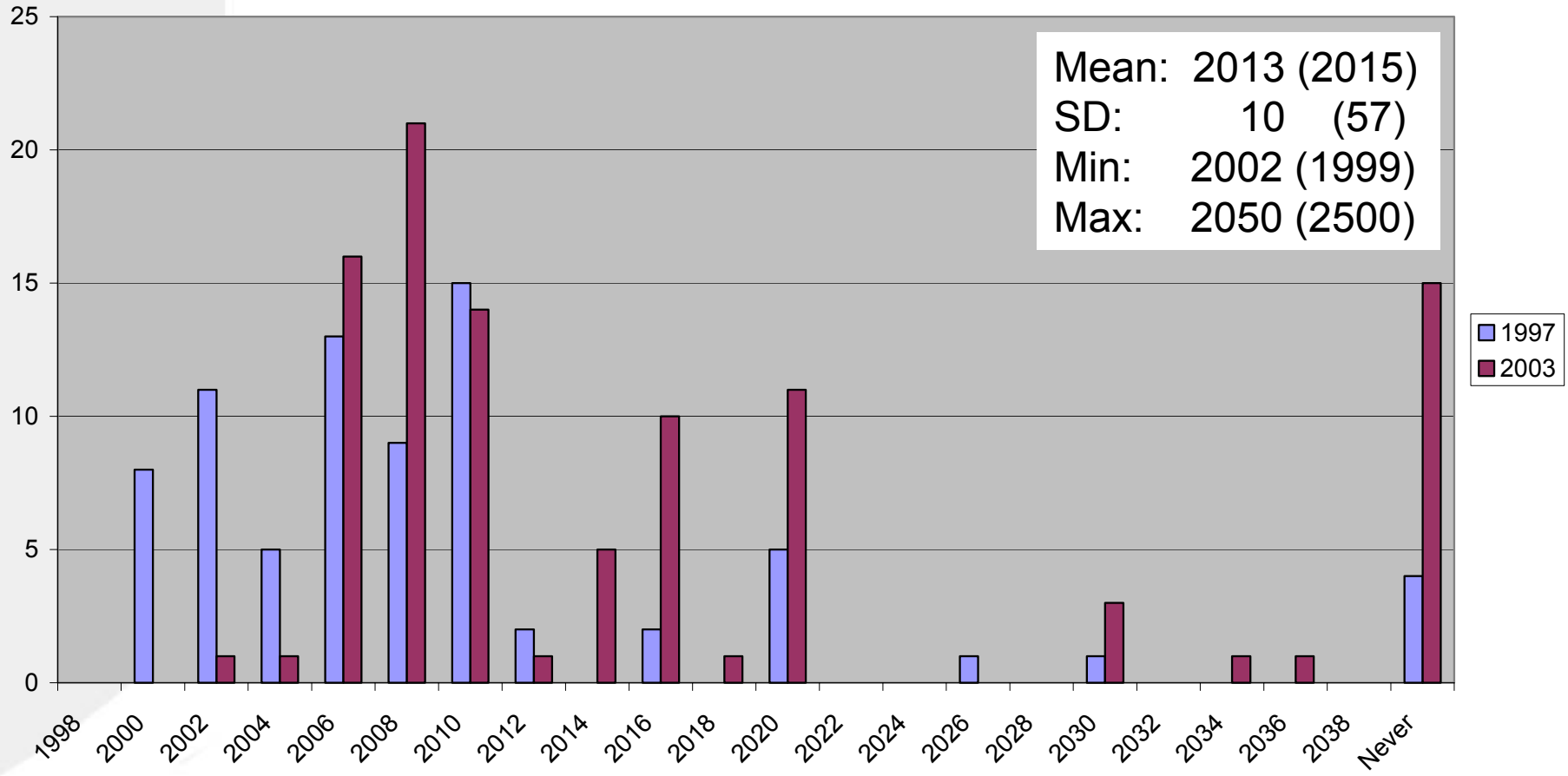


2. Most telephone Interactive Voice Response systems accept speech input (and more than just digits)



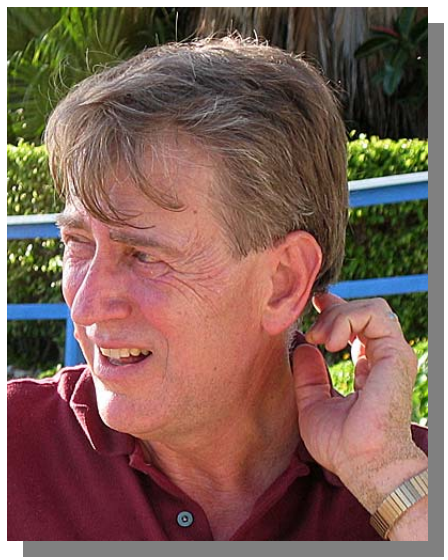
5. Automatic airline reservation by voice over the telephone is the norm.

5. Automatic airline reservation by voice over the telephone is the norm.

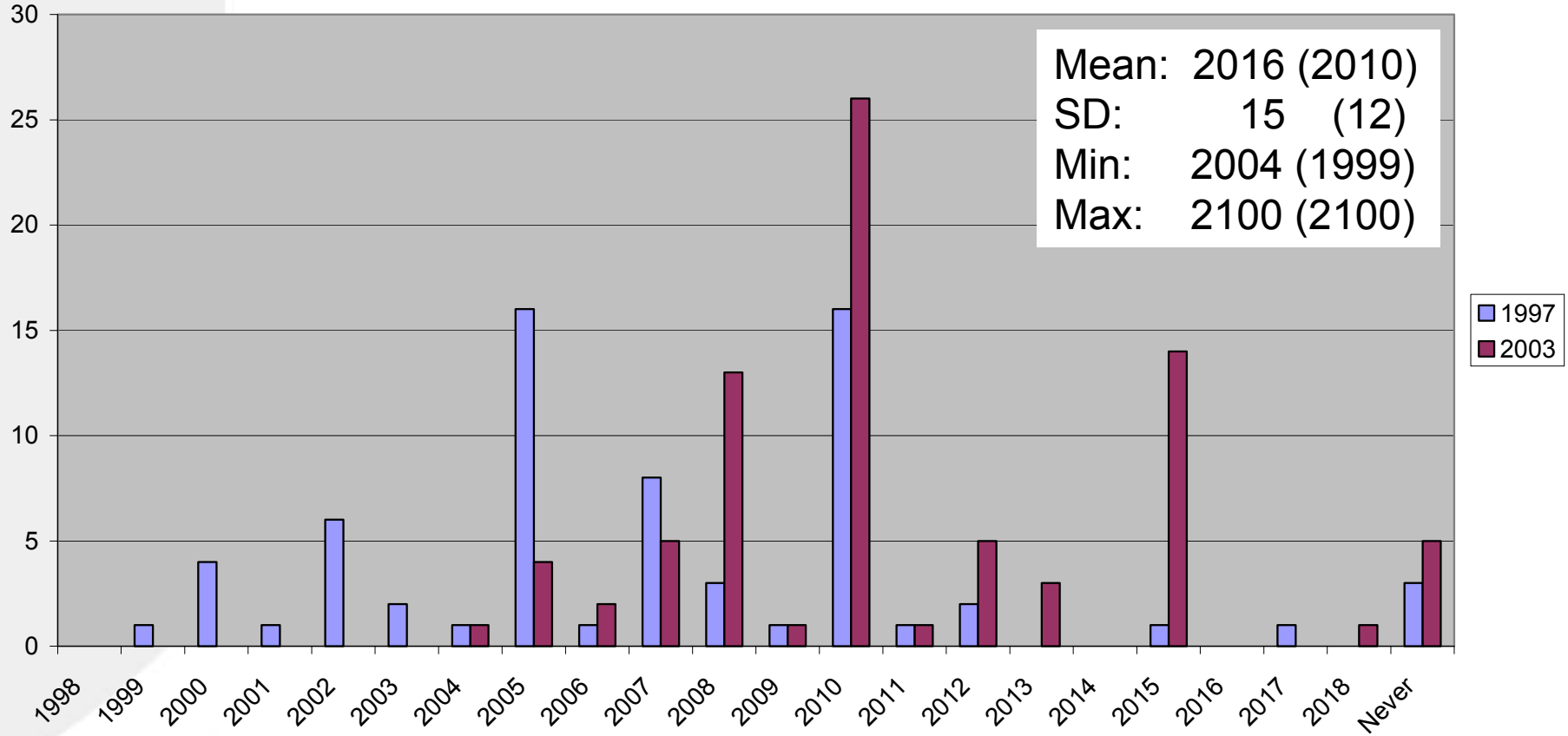


4. Speech recognition is commonly available at home (e.g. interactive TV, control of home appliances and home management systems).

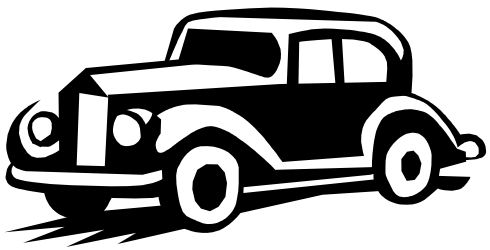
"in your pocket!"

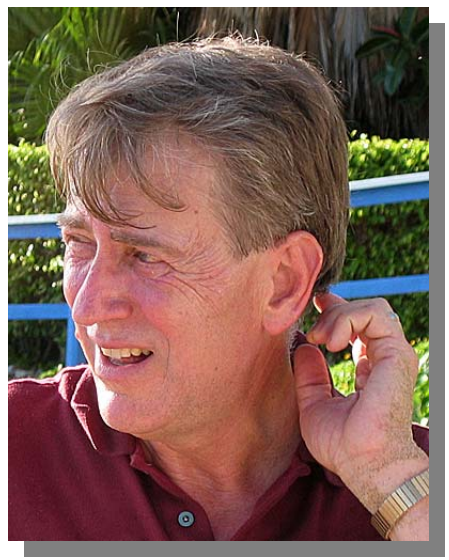


4. Speech recognition is commonly available at home (e.g. interactive TV, control of home appliances and home management systems).

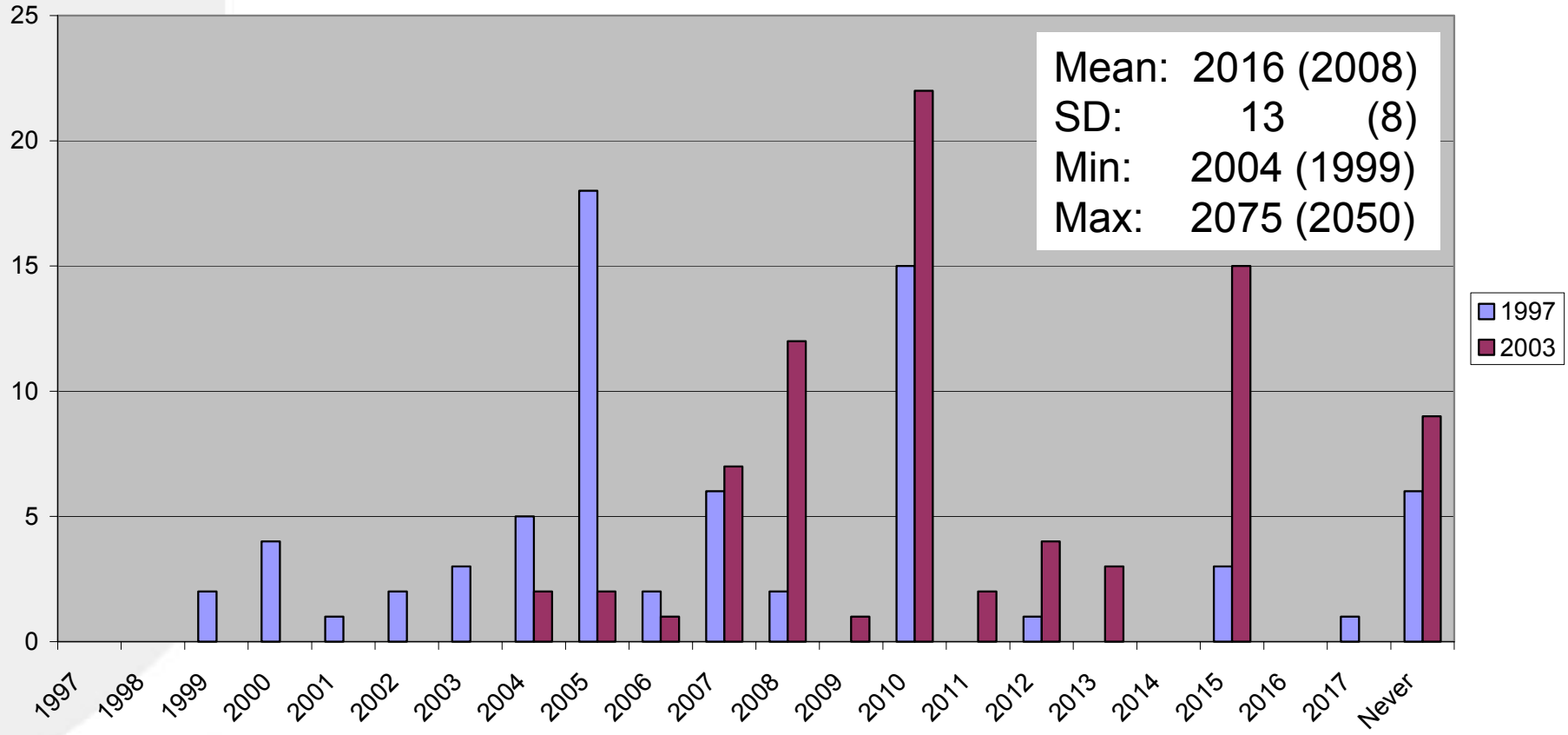


7. Voice-enabled command, control and communication in cars becomes as common as intermittent wiper, power window or power door lock.



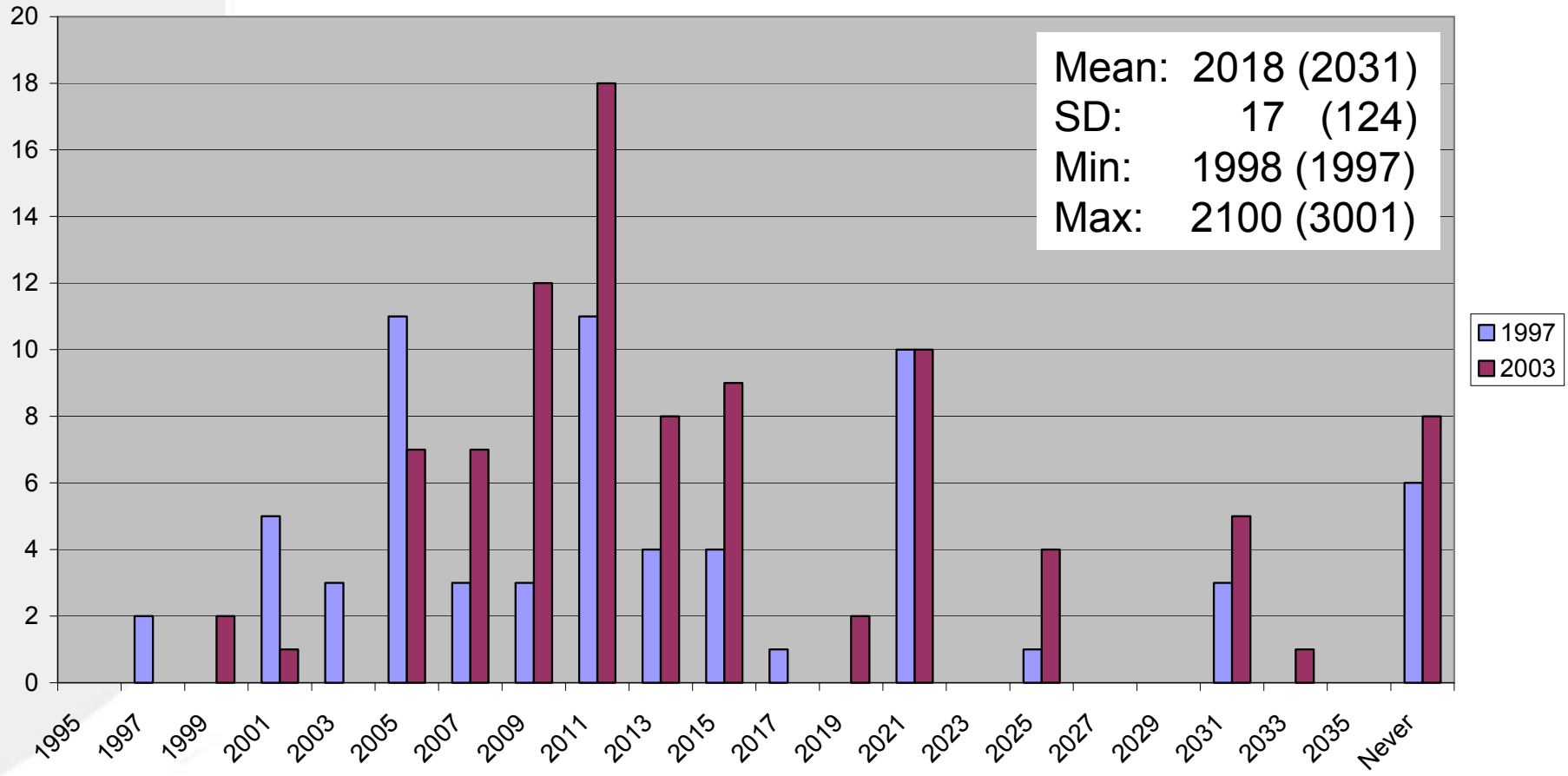


7. Voice-enabled command, control and communication in cars becomes as common as intermittent wiper, power window or power door lock.



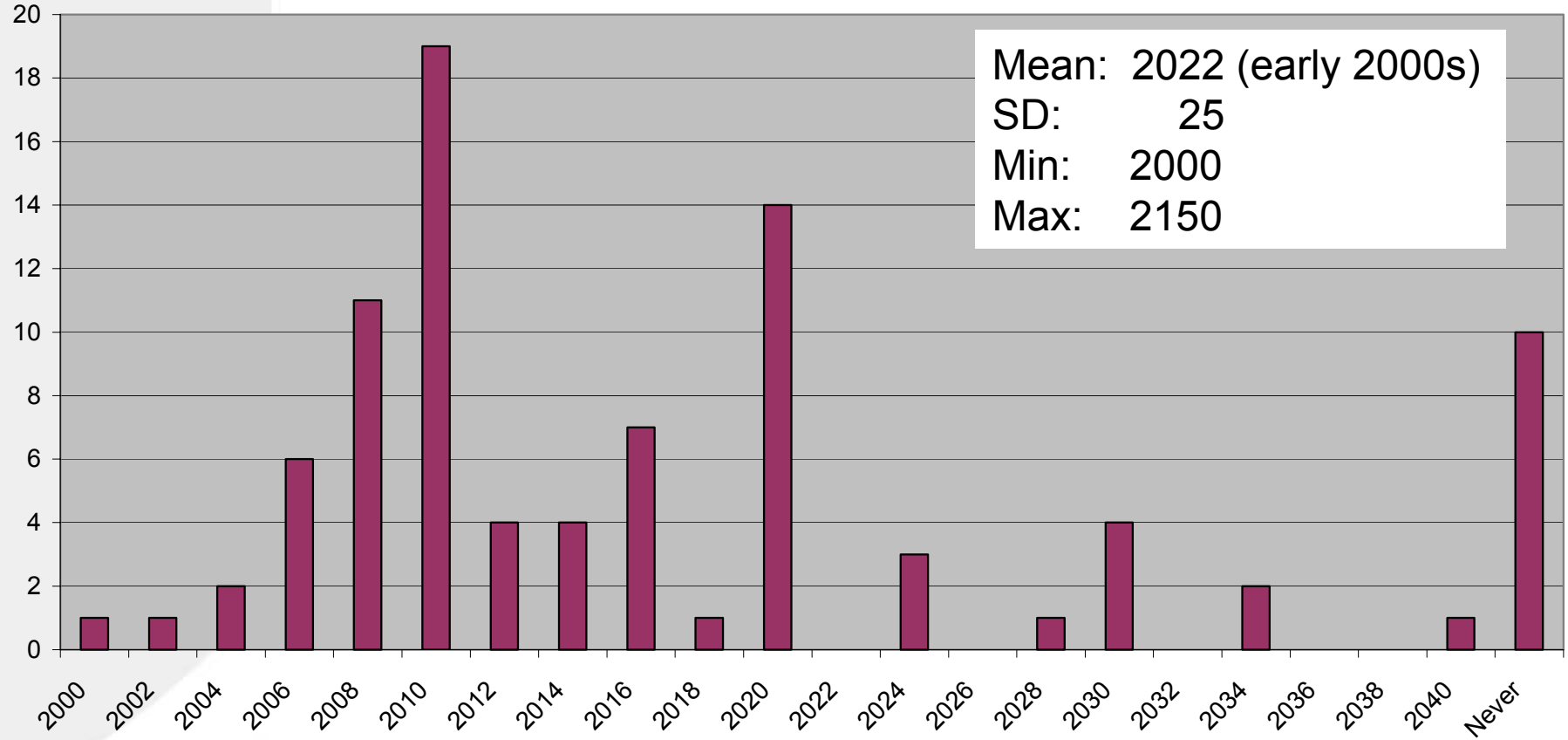
3. TV closed-captioning (subtitling) is automatic and pervasive.

3. TV closed captioning is automatic and pervasive.



15. Telephones are answered by an intelligent answering machine that converses with the calling party to determine the nature and priority of the call.

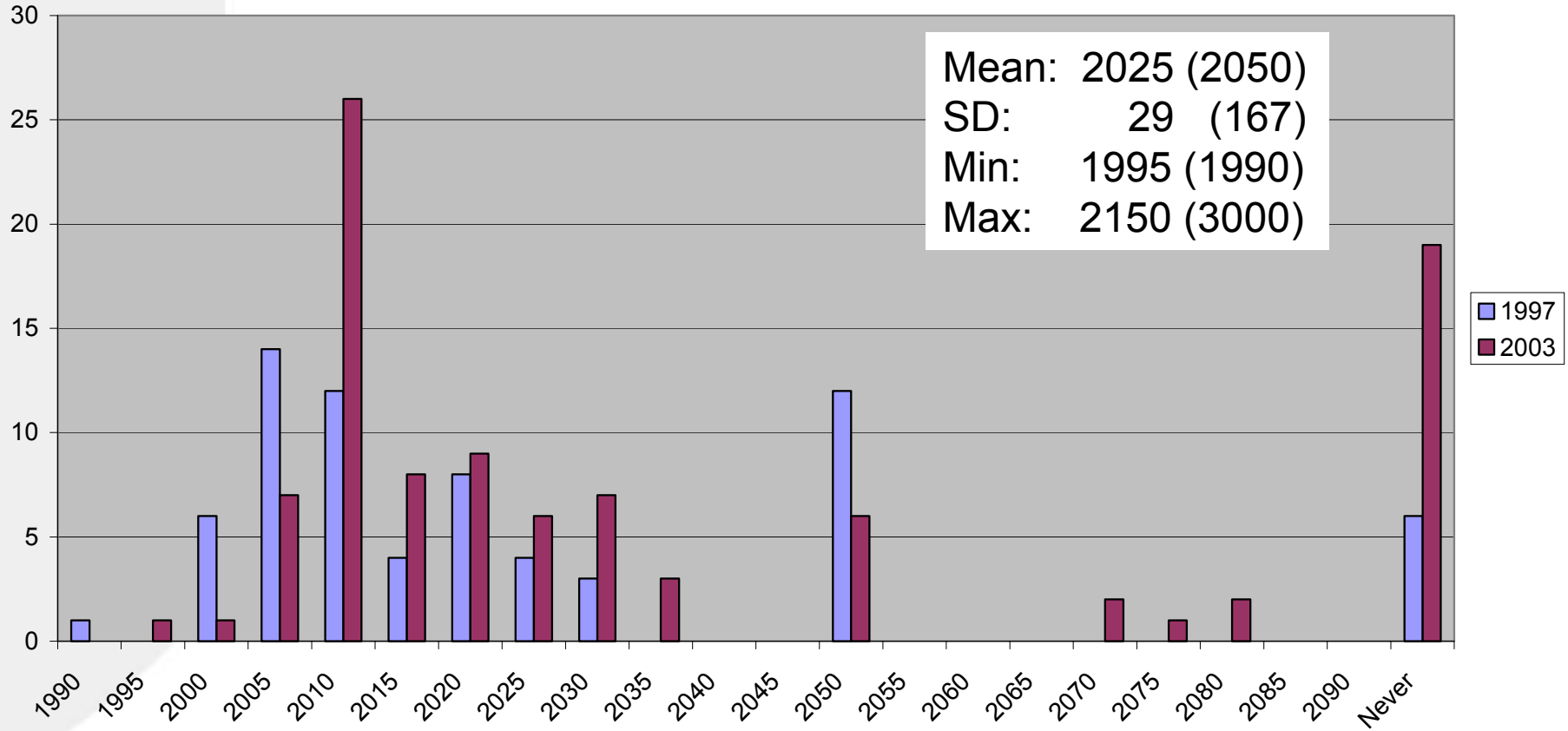
15. Telephones are answered by an intelligent answering machine that converses with the calling party to determine the nature and priority of the call.



11. First legal case in which a recording of a person's voice is thrown out because it cannot be proved whether a computer or a person said it.

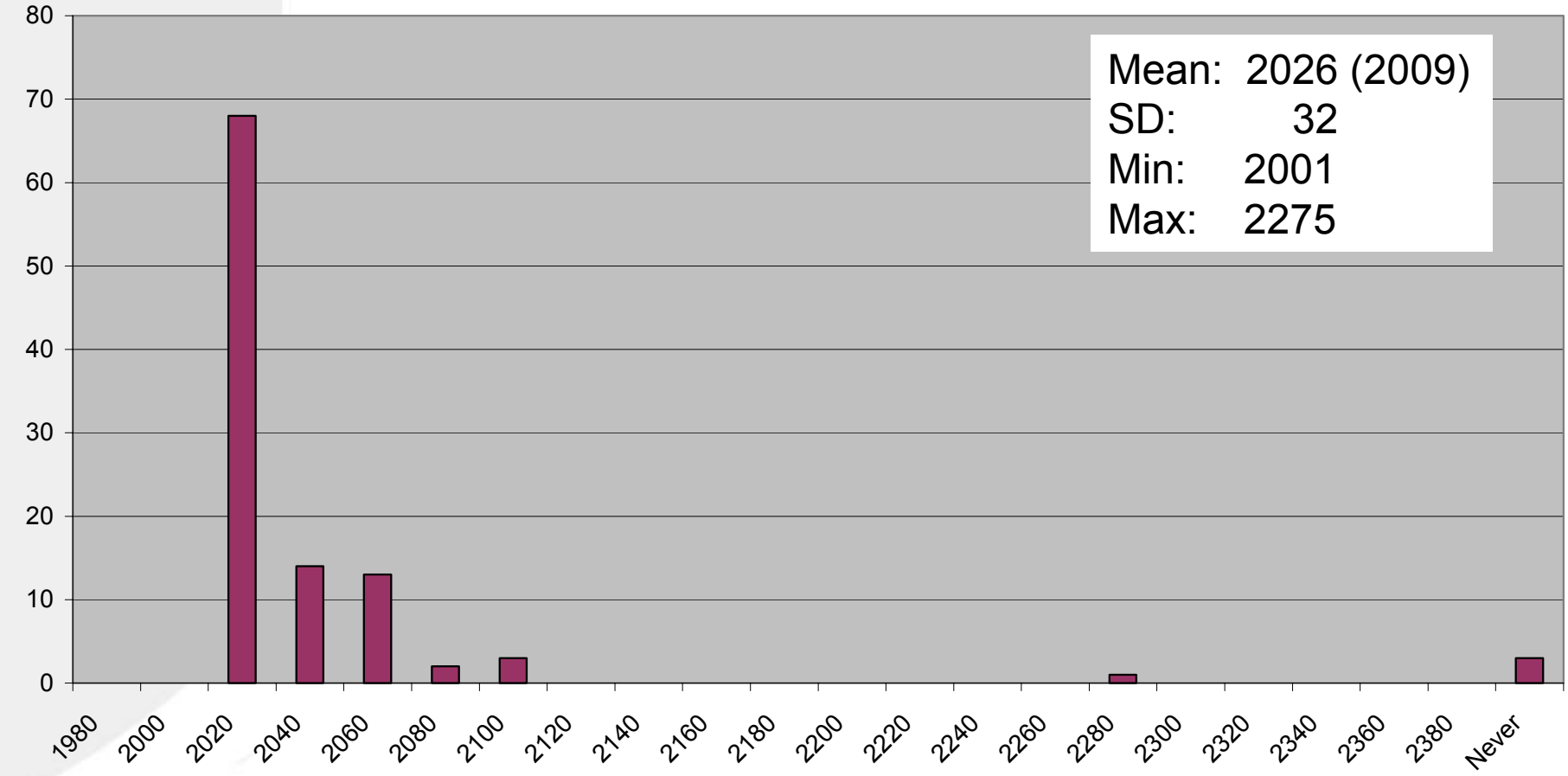
"not evidence in many countries already"

11. First legal case in which a recording of a person's voice is thrown out because it cannot be proved whether a computer or a person said it.



20. Pocket-sized listening machines are commonly available for the hearing impaired.

20. Pocket-sized listening machines are commonly available for the hearing impaired.



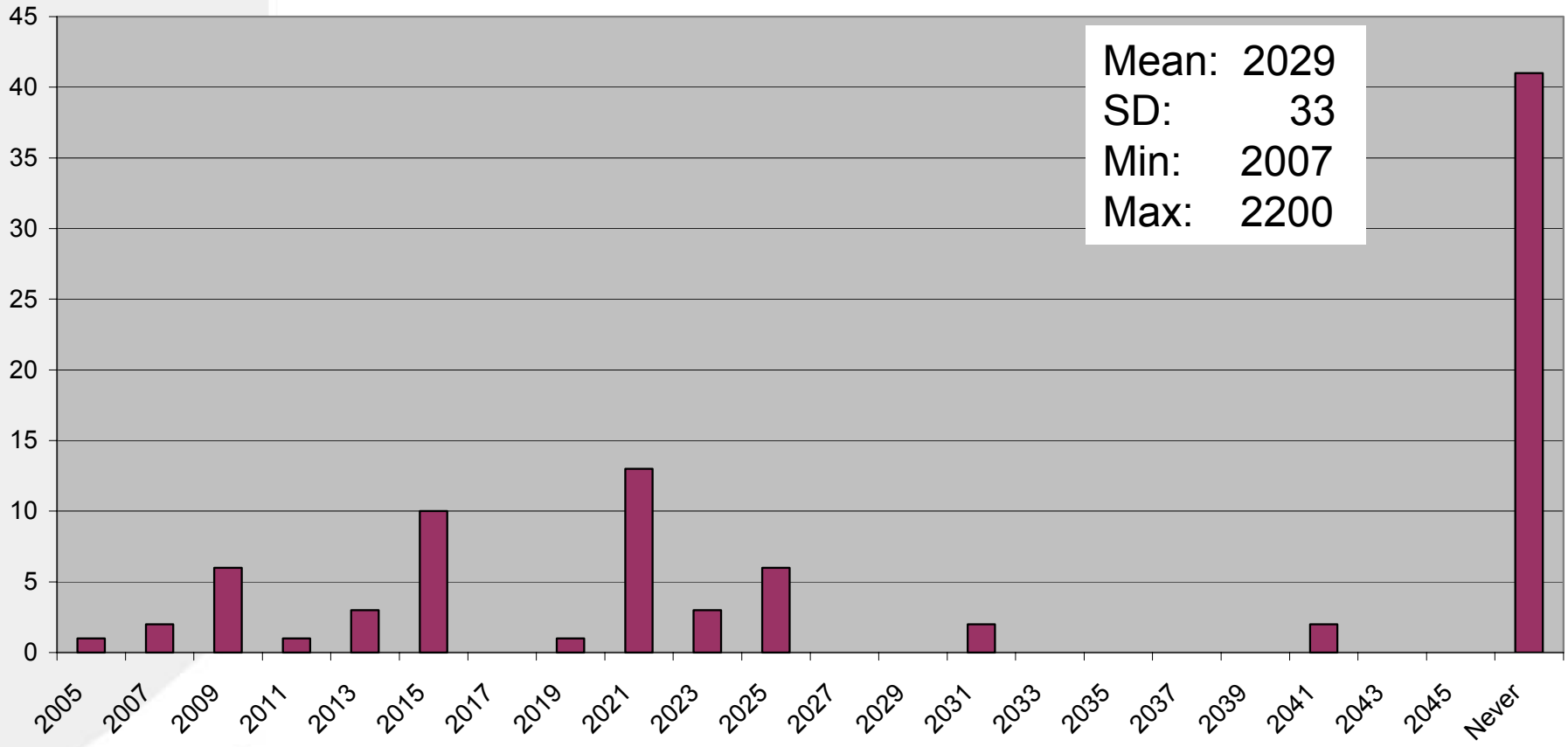
16. The majority of automatic speech recognition systems have completely abandoned the HMM paradigm for acoustic modelling.

"impossible to answer"

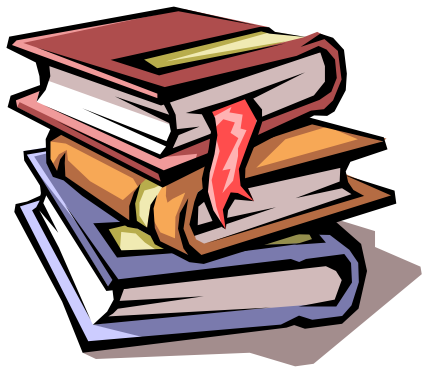
"hope it's soon"

"HMMs are here to stay, but the assumptions will be steadily relaxed"

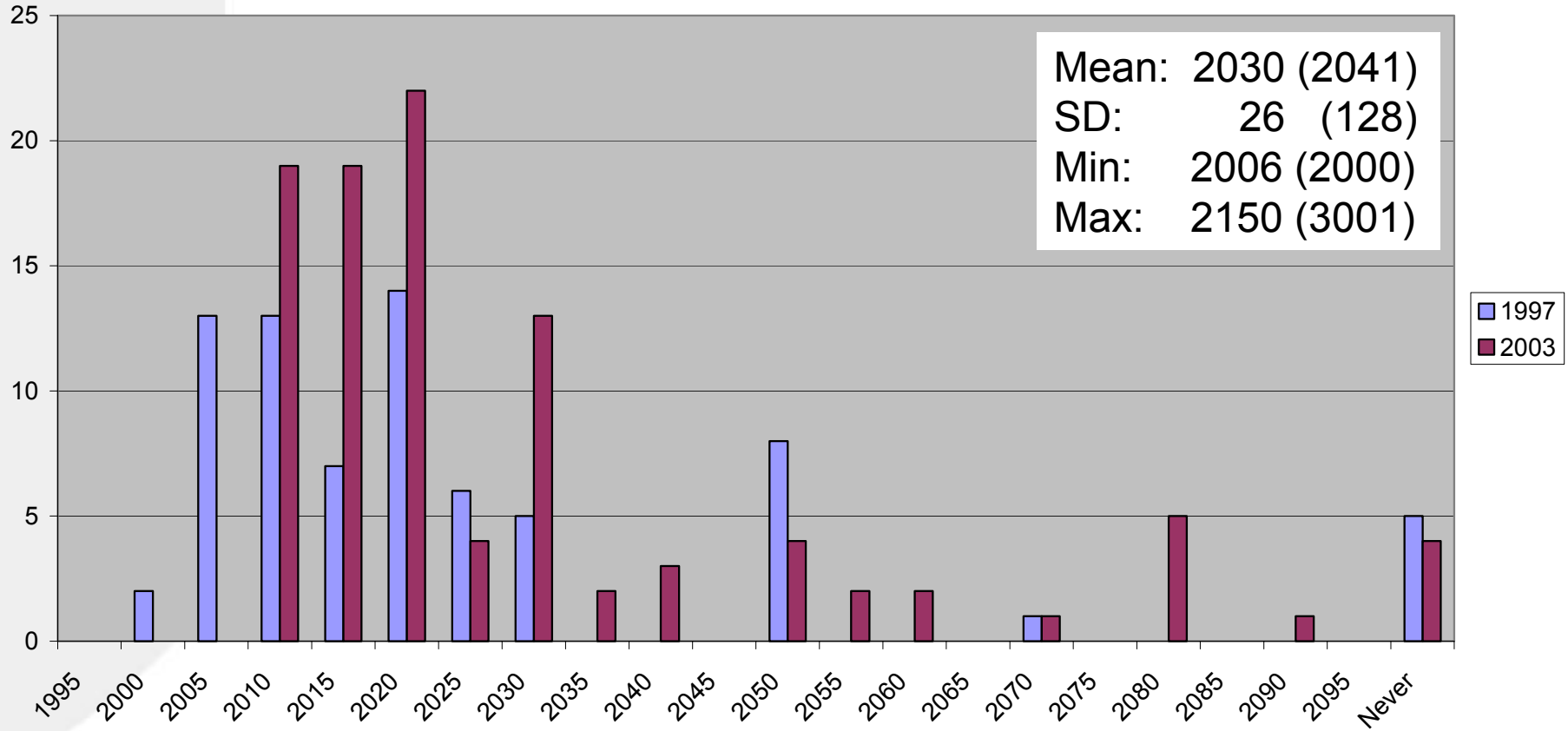
16. The majority of automatic speech recognition systems have completely abandoned the HMM paradigm for acoustic modelling.



10. Public proceedings (e.g. courts, public inquiries, parliament etc.) are transcribed automatically.



10. Public proceeedings (e.g. courts, public inquiries, parliament etc.) are transcribed automatically

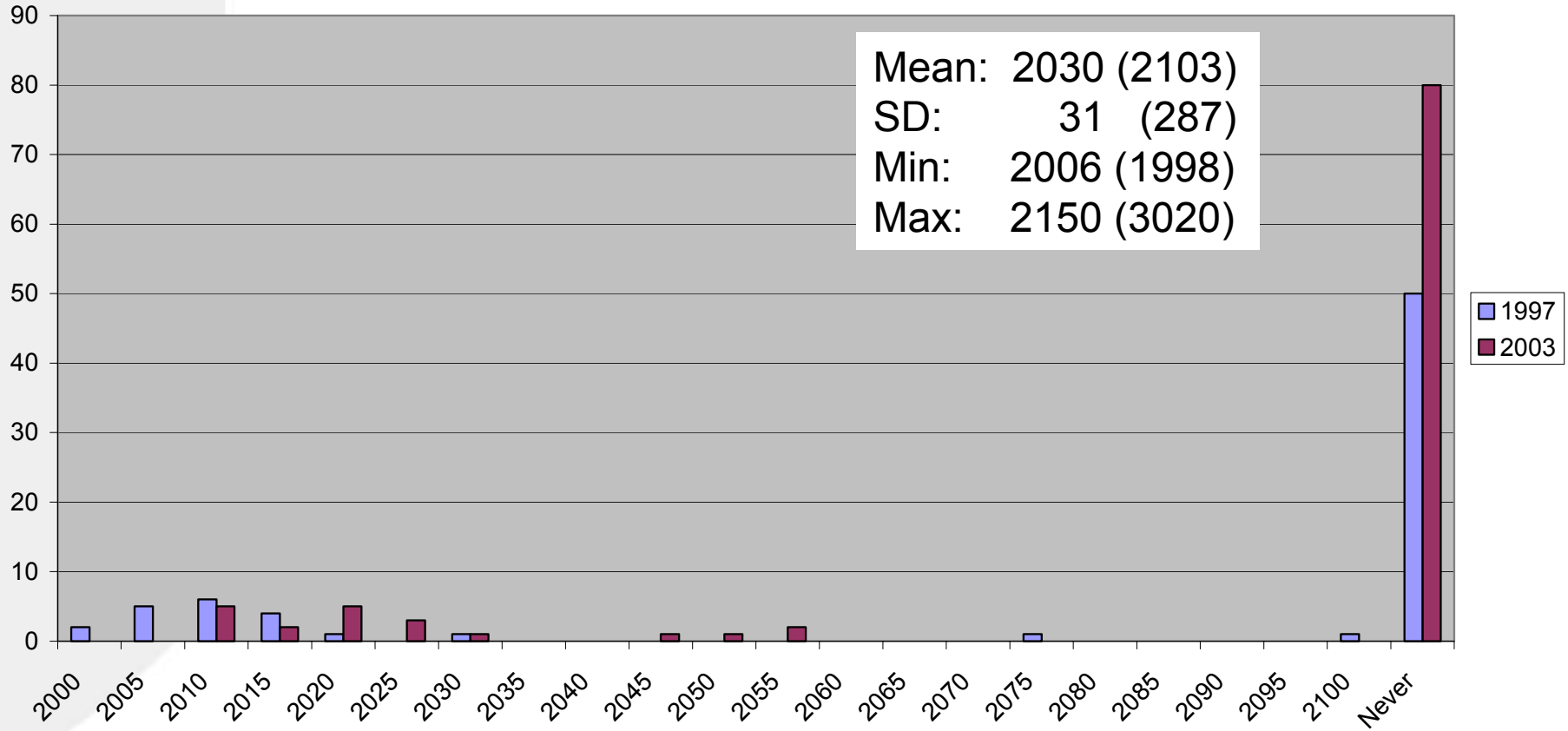


9. A leading cause of time away from work is being hoarse from talking all the time, and people buy keyboards as an alternative to speaking.

"keyboards will always be shipped"

"no problem, there will always be advanced pills available"

9. A leading cause of time away from work is being hoarse from talking all the time, and people buy keyboards as an alternative to speaking.



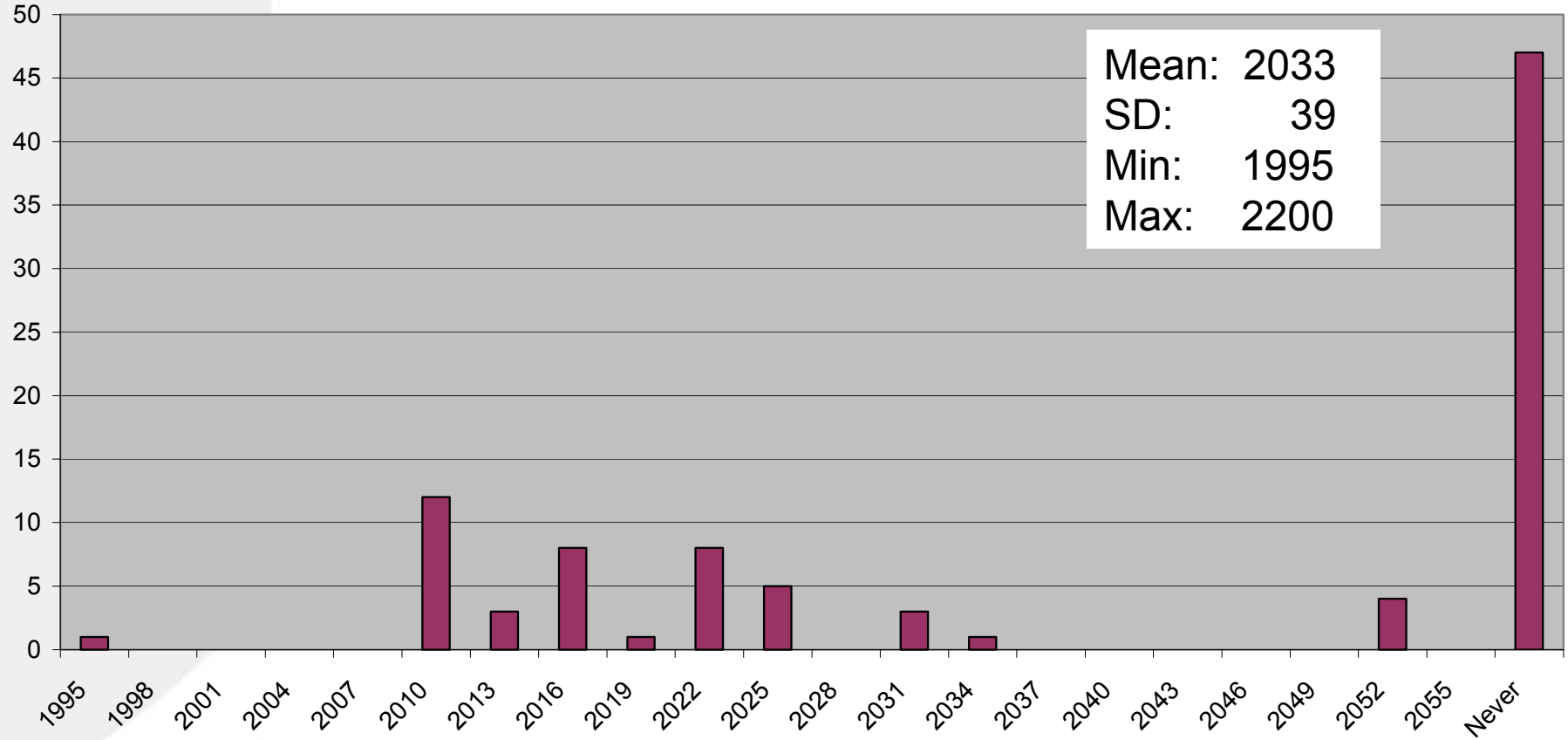
14. The majority of automatic speech recognition systems have completely abandoned the n-grams paradigm for language modelling.

"impossible to answer"

"most deployed systems use CFG anyway"

"hope it's soon"

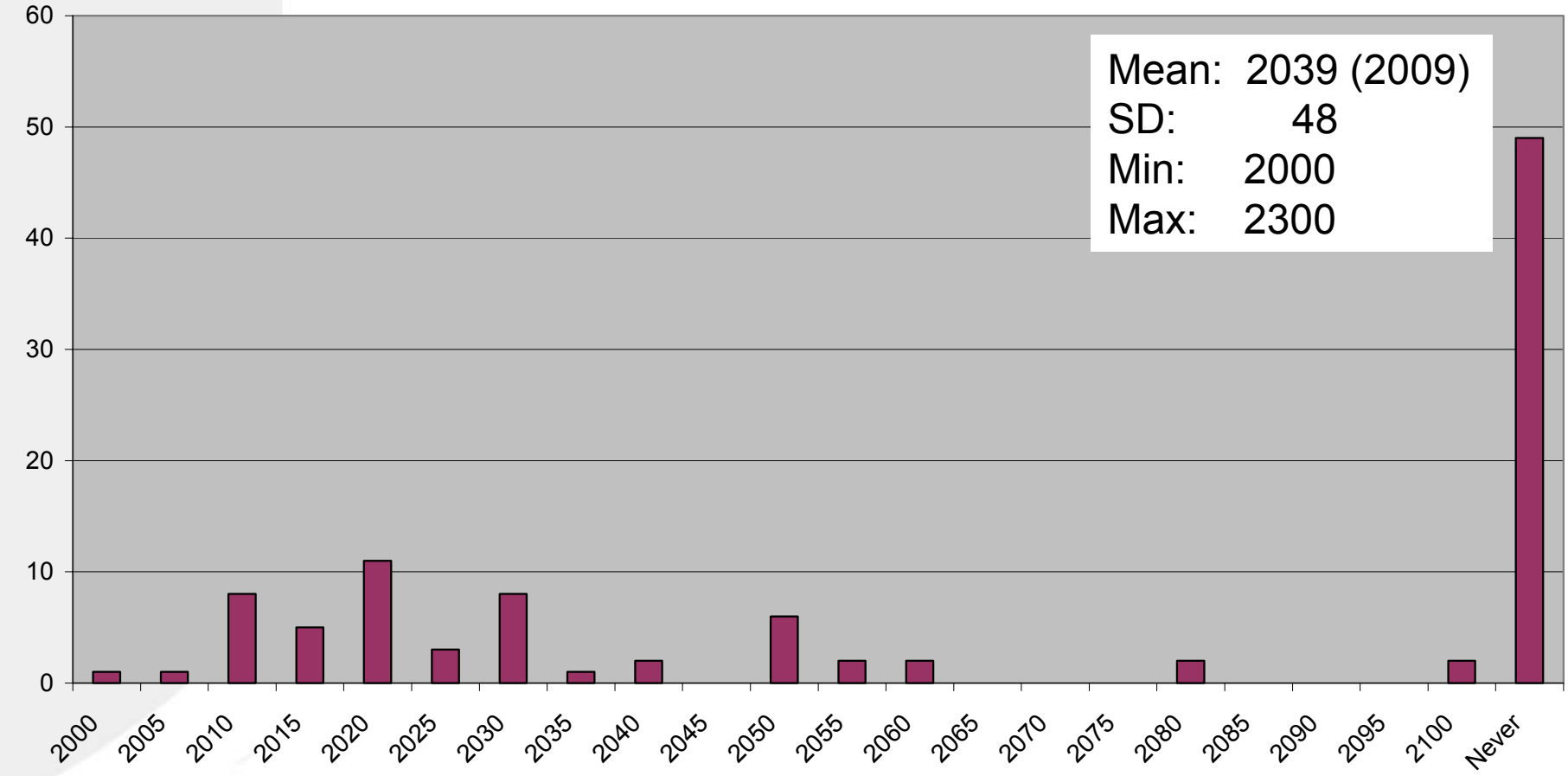
14. The majority of automatic speech recognition systems have completely abandoned the n-grams paradigm for language modelling.



13. The majority of text is created using continuous speech recognition.

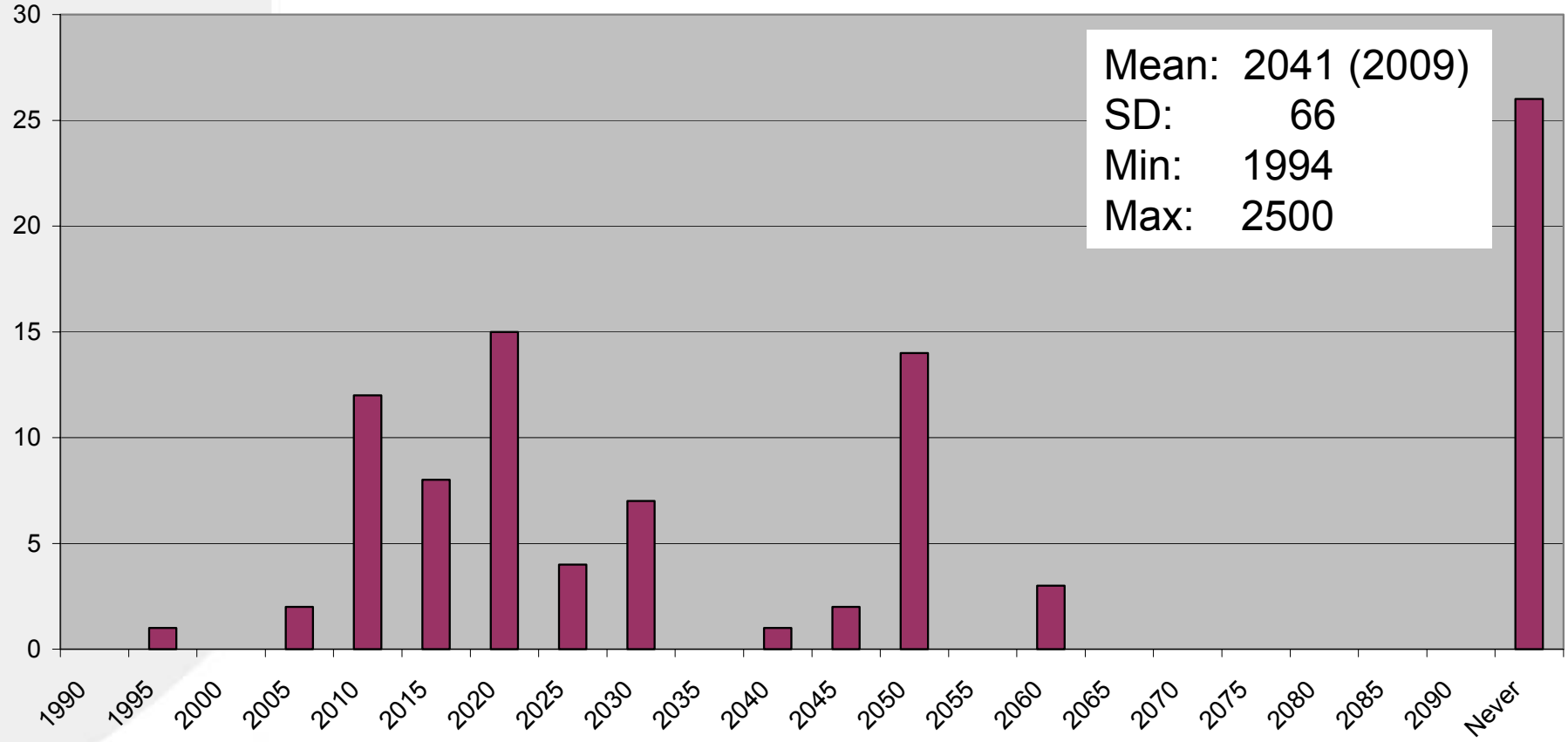
"for authoring
(vs transcription of recordings)"

13. The majority of text is created using continuous speech recognition.



17. Most routine business transactions take place between a human and a virtual personality (including an animated visual presence that looks like a human face).

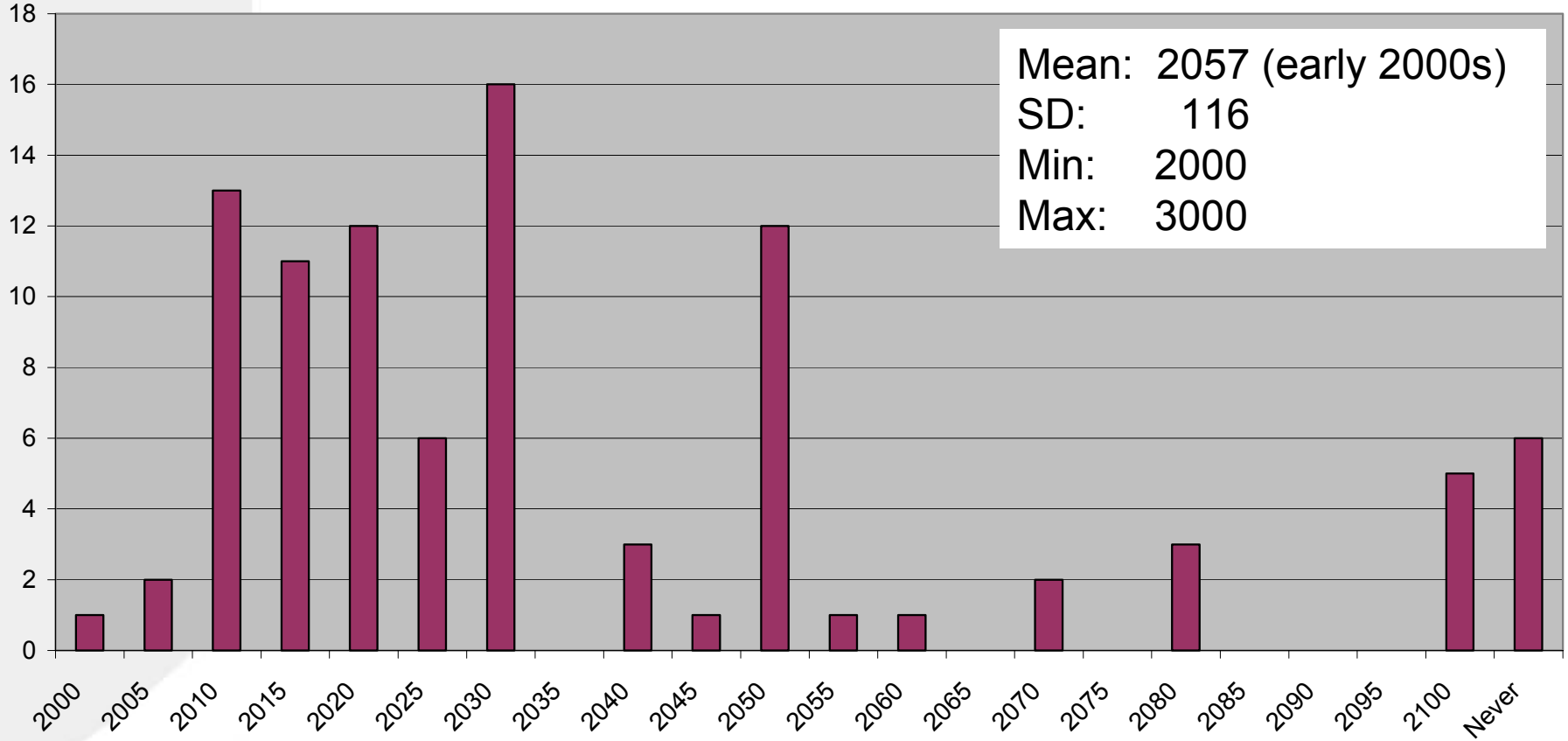
17. Most routine business transactions take place between a human and a virtual personality (including an animated visual presence that looks like a human face).



18. Translating telephones allow two people across the globe to speak to each other even if they do not speak the same language.

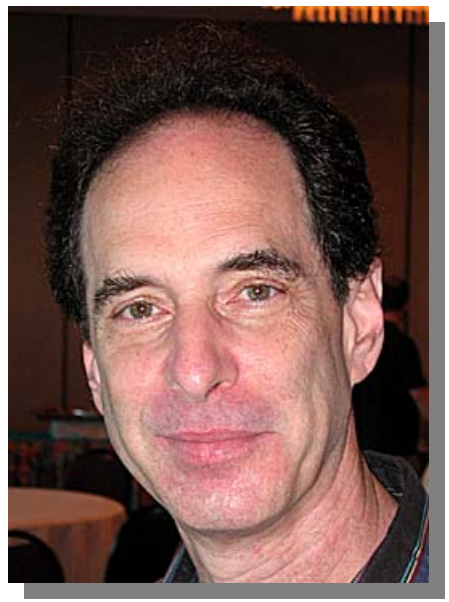
"depends on the task"

18. Translating telephones allow two people across the globe to speak to each other even if they do not speak the same language.

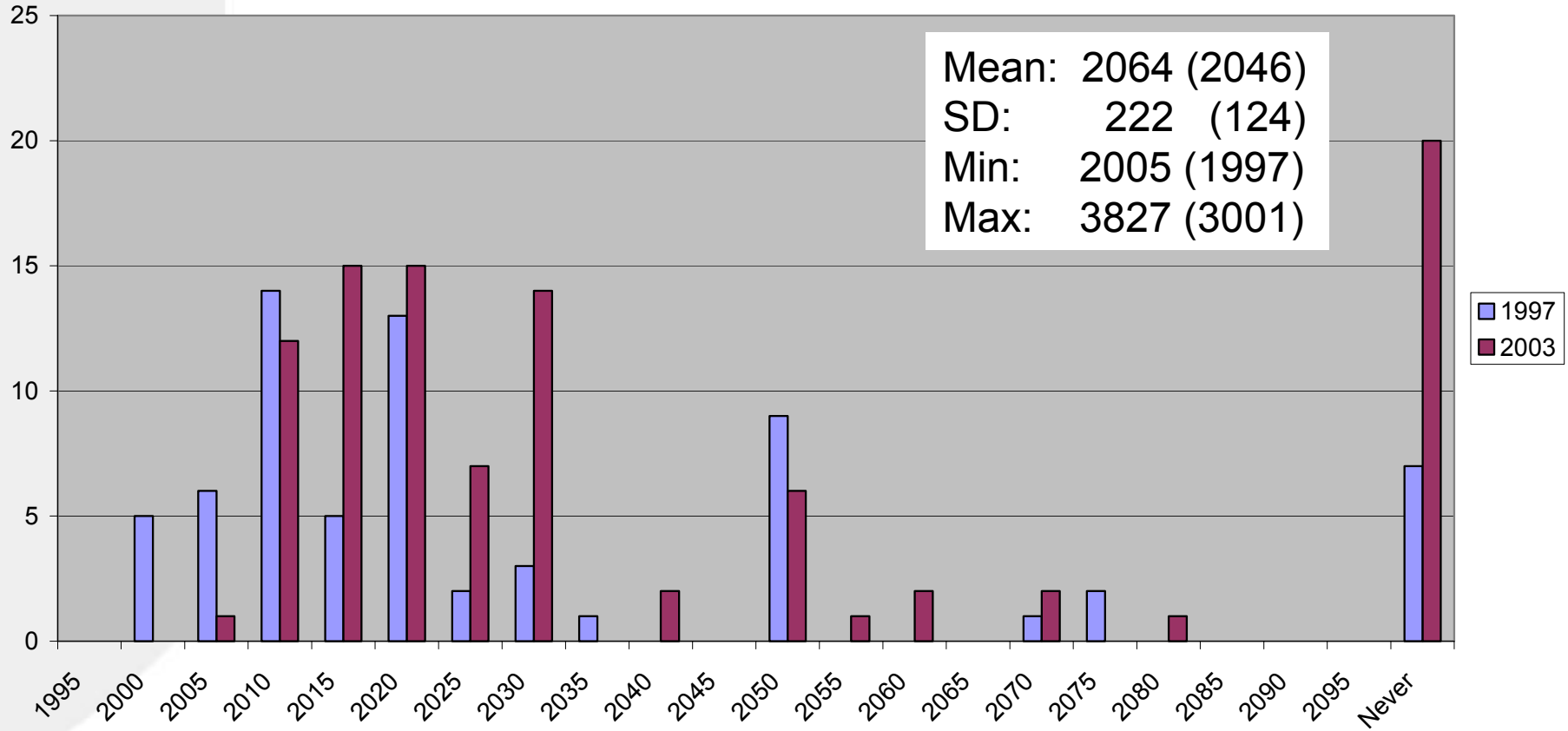


12. Speech recognition accuracy equals that of the average (individual) human transcriber.

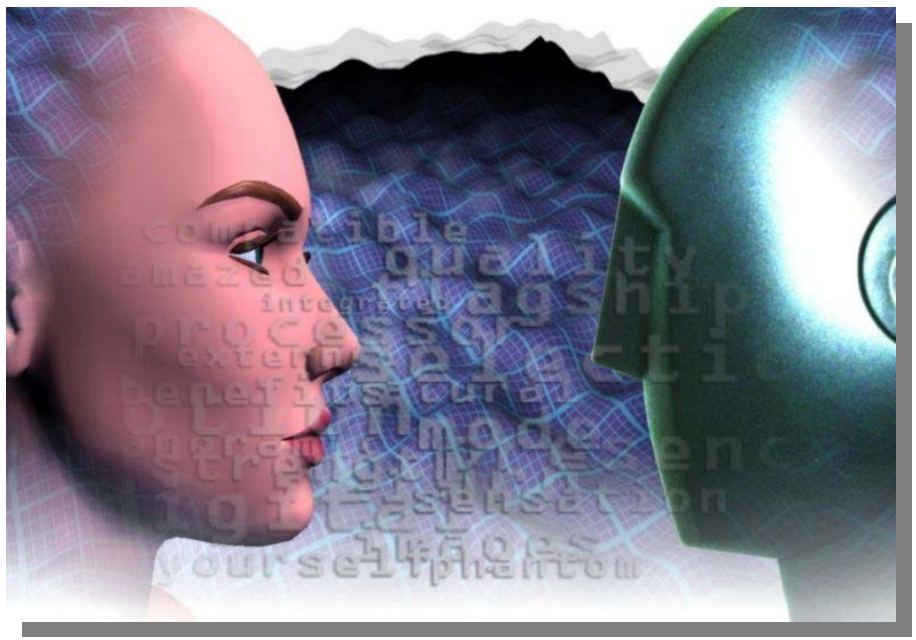
"which task?"



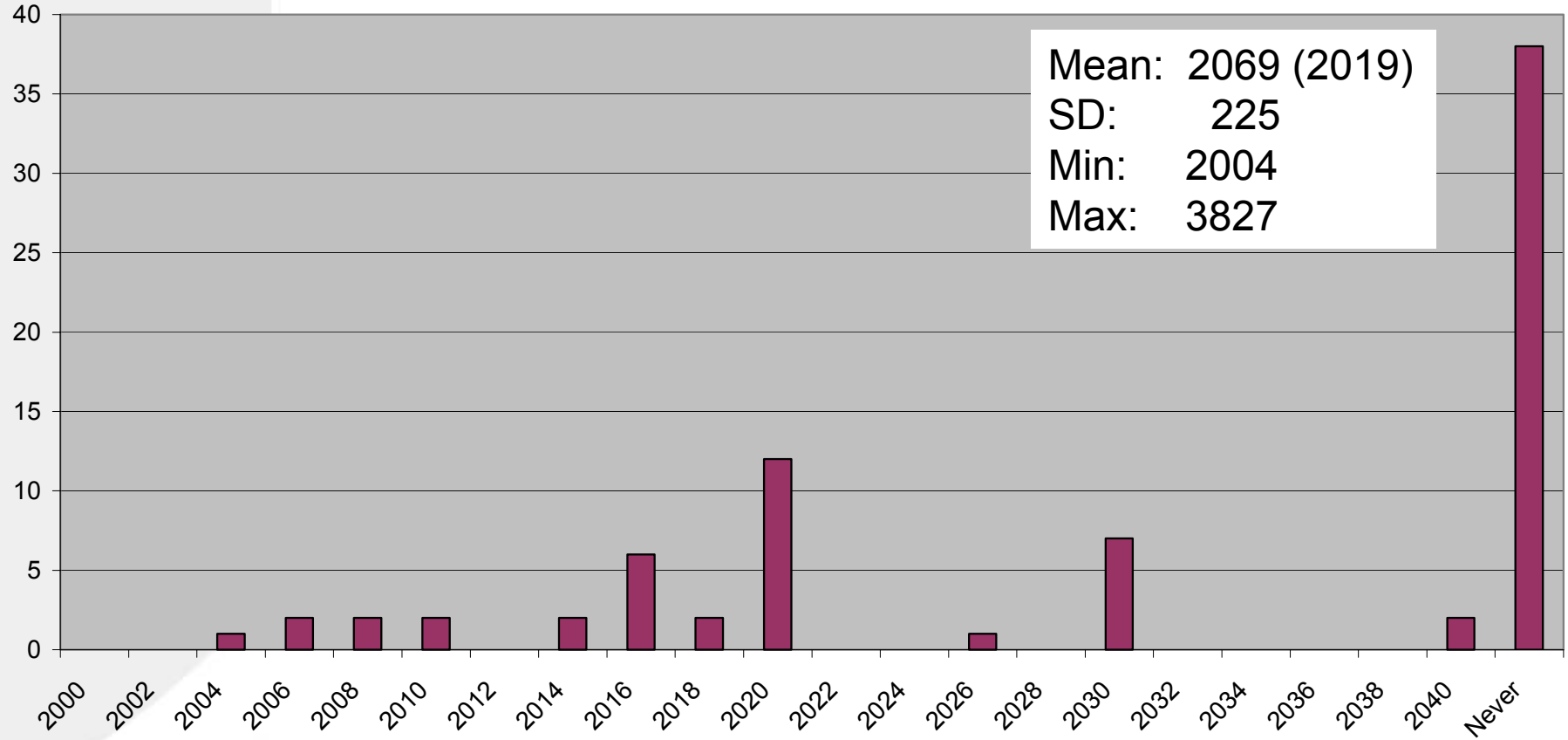
12. Speech recognition accuracy equals that of the average (individual) human transcriber.



19. Most interaction with computing is through gestures and two-way natural-language spoken communication.



19. Most interaction with computing is through gestures and two-way natural-language spoken communication.



aurix

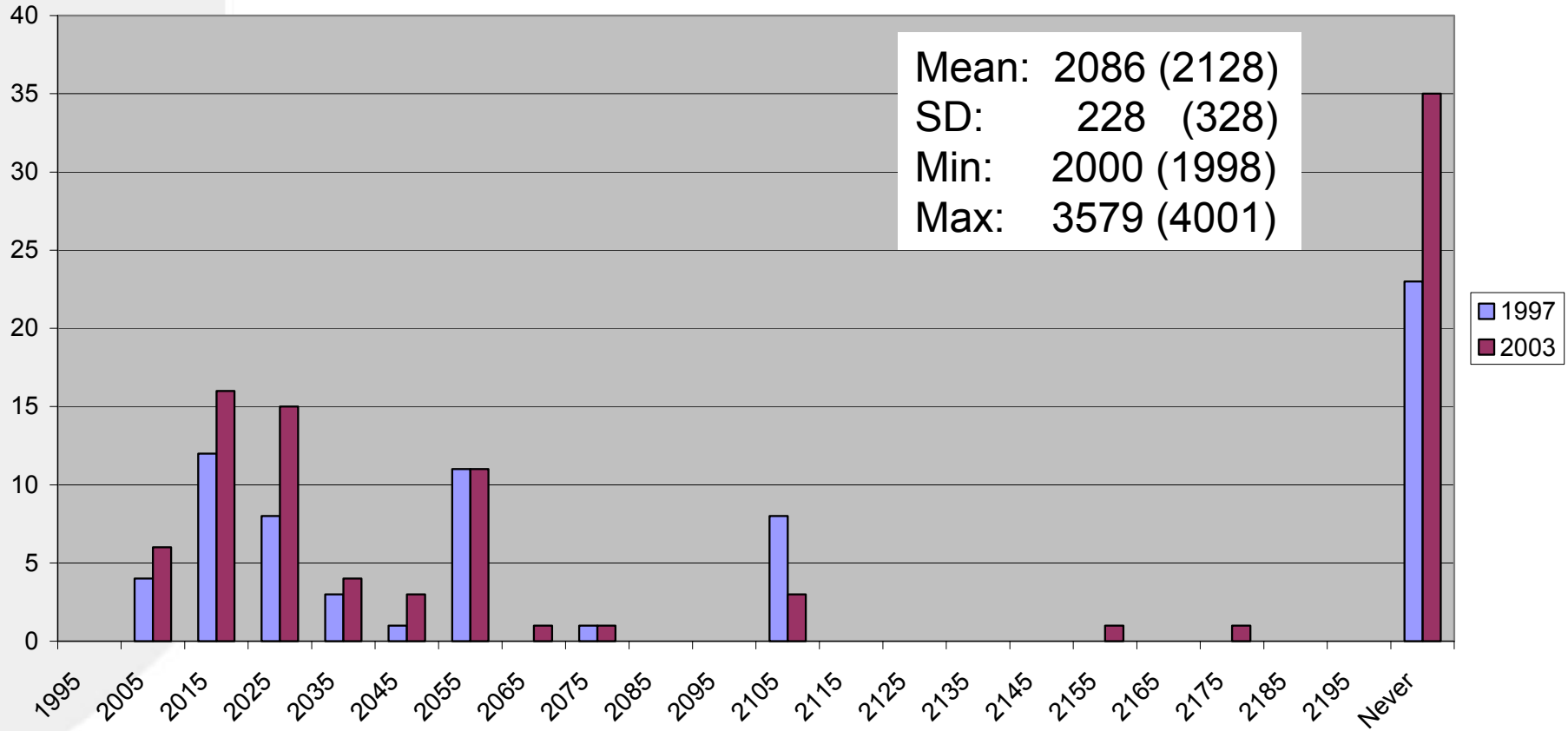
6. It is possible to hold a telephone conversation with an automatic chat-line system for more than 10 minutes without realising it isn't human.

"why would one want that?"

"could happen, but should not"

"hopefully never"

6. It is possible to hold a telephone conversation with an automatic chatline system for more than 10 minutes without realising it isn't human.

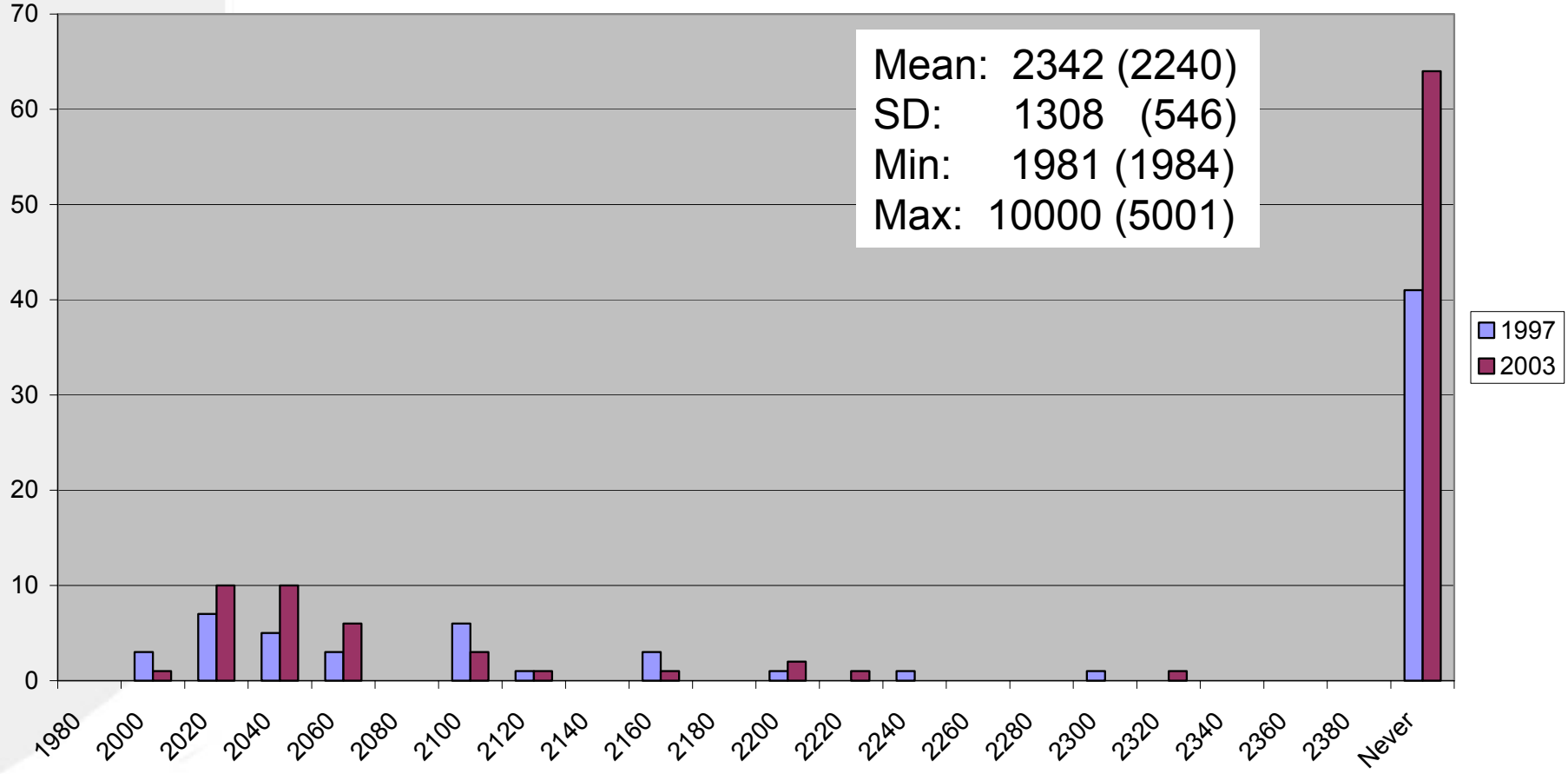


8. No more need for speech research.

“GASP!”

“Ha! Never”

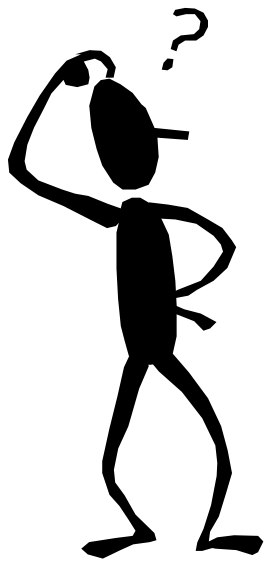
8. No more need for speech research.



Overall Impressions

- High-level of participation – thanks to the 105 who responded
- Remarkably consistent with the 1997 survey
- Strong evidence of the 'Church Effect'
- Neither more optimistic or pessimistic
- More agreement & more realistic
- People less willing to be associated with their opinions than 6 years ago

That’s it Folks !



20/20 Speech Ltd.

Science Park, Malvern, Worcs., WR14 3SZ, UK
Tel: +44 1 684 585101 Fax: +44 1 684 585151

http://www.2020speech.com
http://www.aurix.com